HUMAN BIOLOGY
a record of research

FEBRUARY, 1951

VOL. 23 No. 1

MATHEMATICAL STATISTICAL STUDIES ON THE TUBERCULIN AND HISTOPLASMIN SKIN TESTS

I. OVERLAPPING OF NEGATIVE AND POSITIVE REACTIONS

BY WILLIAM F. TAYLOR
School of Public Health, University of California

SUMMARY

IN interpreting reactions to skin tests there are a number of cases which are indefinite in the sense that the reactions are typical neither of infected persons nor uninfected ones. Palmer and Petersen [1] have presented a mathematical model to aid in studying this which assumes the questionable cases are composed of both infected and uninfected individuals. The Palmer-Petersen model is studied, the parameters are estimated, and a χ^2 test is performed. Some special cases of the model are proposed and tested. The power function of the χ^2 test is used to gain further information about these tests. The derivations and discussion are illustrated by using United States Public Health Service data.

MATERIAL

1. In a paper by Dr. Carroll E. Palmer and O. Strange Petersen [1], reactions to the tuberculin and histoplasmin skin tests are con-

sidered with regard to the occurrence of so-called "doubtful" reactions. Basing their work on the results of skin tests performed on 16,320 student nurses in 11 cities of the United States, Palmer and Petersen develop a statistical model which is intended to explain certain facets of the observations. They assume each population (city) observed is actually divided in two parts: "true positive" reactors, those who are infected with some infecting agent; and "true negative" reactors, those who are not so infected. When an attempt is made to partition a population on the basis of a skin test, however, the members can be segregated only according to definite reactions, doubtful reactions, and no reactions to the test. Palmer and Petersen assume a priori, for practical purposes, that such skin test grouping of a population has real meaning in that no true negative person will be a definite reactor and no true positive a non-reactor. In the case of the tuberculin test it is believed that the definite reactors have received at some previous time a dose of the tubercle bacillus which somehow has caused them to become sensitive to tuberculin. The non-reactors are thought to have escaped such doses, and the doubtful reactors are equivocal in that the interpretation of the reading is uncertain.

The result of testing 16,320 student nurses are given in Tables 1 and 2. From city to city considerable variation is observed in the proportions of persons placed in the three reaction categories. To explain some of this variation, a simple model is set up in terms of the proportions of true positives, true negatives, and the chances of these reacting as definite, doubtful, or non-reactors. It is the primary purpose of the present paper to work with this model in order to estimate the parameters and test the hypotheses involved. The import of the problem can be illustrated by an example. Suppose a city is composed of 50 per cent "true negative" people and 50 per cent "true positive." A skin test, however, calls 90 per cent of the true positives "definite reactors" and 95 per cent of the true negatives "nonreactors." All the others are called "doubtful." This would result in a distribution of reactors as follows:

Definite reactors	45.0%
Doubtful reactors	7.5%
Nonreactors	47.5%
Total	100.0%

Now imagine a second city having 80 per cent true negative people and

TABLE 1

Number of student nurses tested with histoplasmin in specified cities and number and percentage classified as definite, doubtful, and nonreactors

From *Public Health Reports*, Vol. 65, No. 1, pp. 1–32.

CITY	NUMBER OF NURSES TESTED	NUMBER			PERCENTAGE			
		Definite reactors	Doubtful reactors	Non-reactors	Definite reactors	Doubtful reactors	Non-reactors	Total
Kansas City, Missouri	1 131	653	80	398	57.7	7.1	35.2	100
Columbus, Ohio	1 330	740	65	525	55.6	4.9	39.5	100
Kansas City, Kansas	532	241	31	260	45.3	5.8	48.9	100
New Orleans	1 290	273	93	924	21.2	7.2	71.6	100
Baltimore	1 971	434	65	1 472	22.0	3.3	74.7	100
Detroit	1 250	158	21	1 071	12.6	1.7	85.7	100
Los Angeles	1 778	189	65	1 524	10.6	3.7	85.7	100
Philadelphia	2 179	228	71	1 880	10.5	3.2	86.3	100
San Francisco	1 111	78	30	1 003	7.0	2.7	90.3	100
Denver	1 329	88	24	1 217	6.6	1.8	91.6	100
Minneapolis	2 419	101	65	2 253	4.2	2.7	93.1	100
All cities	16 320	3 183	610	12 527	19.5	3.7	76.8	100

TABLE 2

Number of student nurses tested with tuberculin in specified cities and number and percentage classified as definite, questionable, and nonreactors

From *Public Health Reports*, Vol. 65, No. 1, pp. 1–32.

CITY	NUMBER OF NURSES TESTED	NUMBER			PERCENTAGE			
		Definite reactors	Questionable reactors	Non-reactors	Definite reactors	Questionable reactors	Non-reactors	Total
Kansas City, Missouri	1 131	123	61	947	10.9	5.4	83.7	100
Columbus, Ohio	1 330	151	44	1 135	11.4	3.3	85.3	100
Kansas City, Kansas	532	73	41	418	13.7	7.7	78.6	100
New Orleans	1 290	171	185	934	13.3	14.3	72.4	100
Baltimore	1 971	353	112	1 506	17.9	5.7	76.4	100
Detroit	1 250	174	52	1 024	13.9	4.2	81.9	100
Los Angeles	1 778	333	88	1 357	18.7	5.0	76.3	100
Philadelphia	2 179	497	98	1 584	22.8	4.5	72.7	100
San Francisco	1 111	207	44	860	18.6	4.0	77.4	100
Denver	1 329	250	68	1 011	18.8	5.1	76.1	100
Minneapolis	2 419	192	74	2 153	7.9	3.1	89.0	100
All cities	16 320	2 524	867	12 929	15.5	5.3	79.2	100

20 per cent true positives. Then with the same 90 per cent and 95 per cent as above the reactors would be distributed as

Definite reactors	18%
Doubtful reactors	6%
Nonreactors	76%

Thus a certain amount of variation can be accounted for by a very simple model which recognizes the doubtful group as composed of both positive and negative persons.

It is now assumed that the group of student nurses of a given city is made up of individuals who are either "true positive" or "true negative" persons. Secondly it is assumed that the chance of a true positive person being called a definite reactor is the same no matter what the city. Similarly, the chance of a true negative being called a nonreactor is constant. Hence, the variation in the distribution of reactors is due to variation in the proportions of true positives and true negatives. Statistically the problem is to fit this model to the observations and then try to determine if the fit is adequate.

MATHEMATICAL MODEL

2. In setting up this problem on a more formal basis one must first see what is available in the data and then connect the proposed model with the observations. From each of L cities ($L = 11$) there is a sequence of observations. Suppose that n_i persons are tested in the i-th city, $i = 1, \cdots, L$. Denote the relative frequencies of the outcomes as follows:

x_i = proportion of definite reactions in i-th city,
y_i = " " no " " " " ",
z_i = " " doubtful " " " " ",

where the frequencies are $n_i x_i$, $n_i y_i$, and $n_i z_i$, so that the identity

$$x_i + y_i + z_i = 1$$

holds for all i.

Let

p_i = proportion of true positives in the i-th city,
$1 - p_i$ = proportion of true negatives in the i-th city,
α = probability of getting a doubtful reaction from a true positive, and
β = probability of getting a doubtful reaction from a true negative.

Then the probability of various reactions among those in the i-th city are ξ_i, η_i, and ζ_i, and may be given by

(1) $P\{\text{definite reaction in } i\text{-th city}\} = \xi_i = p_i(1-\alpha)$,

(2) $P\{\text{no reaction in } i\text{-th city}\} = \eta_i = (1-p_i)(1-\beta)$,

(3) $P\{\text{doubtful reaction in } i\text{-th city}\} = \zeta_i = p_i\alpha + (1-p_i)\beta$.

The expected frequencies of various reactors are $n_i\xi_i$, $n_i\eta_i$, $n_i\zeta_i$, with $\xi_i + \eta_i + \zeta_i = 1$ for all i. Since $\zeta_i = 1 - \xi_i - \eta_i$, one need work only with ξ_i and η_i.

The task now is to estimate $\alpha, \beta, p_i, \xi_i, \eta_i, \zeta_i$ $(i = 1, \cdots, L)$; i. e., to fit the model to the data. One may consider this problem as involving sequences of independent trials. A trial consists of the performing of a skin test. The data used are records of the first tuberculin or the first histoplasmin test given to student nurses as they entered training. The outcomes of such tests are treated as being mutually independent. Each observation falls in one of three possible groups: definite reaction, doubtful reaction, or no reaction. (Actually the observations were sorted into 7 categories, but these have been grouped to give the three above.) Thus, for both tuberculin and histoplasmin skin tests, there are L sequences of independent observations, those in each sequence having three possible outcomes. The independent probabilities involved are ξ_i and η_i $(i = 1, \cdots, L)$ and these are functions of $L+2$ unknown independent parameters α, β, $p_1, \cdots, p_L$. Since there are $2L$ independent probabilities, $(L > 2)$, best asymptotically normal (BAN) estimates can be found for these parameters. The procedure followed is the one developed by Neyman in [2] and illustrated by Lim in [3]. See Section 9 for a discussion of estimation.

The method of estimation carried out below is a variation of the least squares technique and is concerned with finding values for the unknown parameters which minimize the expression

$$(4) \qquad \chi^2 = \sum_{i=1}^{L} n_i \left[\frac{(x_i - \xi_i)^2}{x_i} + \frac{(y_i - \eta_i)^2}{y_i} + \frac{(z_i - \zeta_i)^2}{z_i} \right],$$

where ξ_i, η_i, ζ_i are given by (1), (2), and (3). Such minimizing values are BAN estimates of the unknown parameters. An equivalent procedure is obtained if the parameters, $\alpha, \beta, p_1, \cdots, p_L$ are eliminated from equations (1), (2), and (3). This results in several conditions being imposed on the ξ's and η's which are expressible as functions of

the ξ's and η's equated to zero, $F_k(\xi, \eta) = 0$. χ^2 can now be minimized with respect to the ξ's and η's, subject to the conditions $F_k(\xi, \eta) = 0$, and the resulting minimizing values are BAN estimates of the ξ's and η's. These provide the same estimates of α, β, $p_1, \cdots, p_L$ as in the first method above.

It is often found that both of the methods above are practically impossible to carry out due to algebraic difficulties. To combat this a third estimation technique has been evolved from the second by Neyman [2]. The set of conditions, $F_k(\xi, \eta) = 0$, often the source of trouble, is put in a "reduced" form, i. e., each function, $F_k(\xi, \eta)$, is expanded in a Taylor's expansion about the point observed, $\xi_i = x_i$, $\eta_i = y_i$, $\zeta_i = z_i$; $i = 1, \cdots, L$. Only the constant and linear terms, $F_k^*(\xi, \eta)$, of this expansion are retained. Upon being equated to 0 they are called the "reduced" conditions on the ξ's and η's, and the obtaining of estimates proceeds as indicated above, replacing F_k by F_k^*. That these are BAN estimates is proven in [2]. The application of this last method follows.

Since

$$\xi_i = p_i(1 - \alpha), \tag{1}$$

$$\eta_i = (1 - p_i)(1 - \beta), \text{ and} \tag{2}$$

$$\zeta_i = p_i\alpha + (1 - p_i)\beta, \tag{3}$$

it can be shown easily that the conditions on the probabilities can be written as the $L - 2$ equations

$$F_k(\xi, \eta) = (\xi_2 - \xi_1)\eta_k + (\eta_1 - \eta_2)\xi_k - (\eta_1\xi_2 - \eta_2\xi_1) = 0, \quad k = 3, \cdots, L. \tag{5}$$

These are obtained by putting $p_i = \xi_i/(1 - \alpha)$ and then eliminating α and β from the expressions for η_1 and η_2. The parameters are expressed in terms of the ξ's and η's as

$$p_i = \frac{\xi_i(\eta_1 - \eta_2)}{\eta_1\xi_2 - \eta_2\xi_1}, \tag{6}$$

$$\alpha = 1 - \frac{\eta_1\xi_2 - \eta_2\xi_1}{\eta_1 - \eta_2}, \tag{7}$$

$$\beta = 1 - \frac{\eta_1\xi_2 - \eta_2\xi_1}{\xi_2 - \xi_1}. \tag{8}$$

Expanding $F_k(\xi, \eta)$ in a Taylor's expansion and keeping only the linear terms gives the reduced conditions

$$(9)\quad F_k^*(\xi,\eta) = (x_2-x_1)y_k + (y_1-y_2)x_k - (y_1x_2-y_2x_1) + (y_2-y_1)(\xi_1-x_1) + (y_k-y_1)(\xi_2-x_2) + (y_1-y_2)(\xi_k-x_k) + (x_k-x_2)(\eta_1-y_1) + (x_1-x_2)(\eta_2-y_2) + (x_2-x_1)(\eta_k-y_k) = 0.$$

To find the values of the ξ's and η's which minimize χ^2 subject to the restraints $F_k^* = 0$ one uses the method of Lagrange multipliers.

Let

$$(10)\qquad \Phi = \chi^2 - 2\sum_{k=3}^{L}\lambda_k F_k^*(\xi,\eta)$$

where χ^2 is given by (4) and the λ_k's are constants. Then in order to minimize χ^2, differentiate Φ partially with respect to the ξ's and η's. The derivatives can be written as:

$$(11)\quad \left(\frac{n_1}{x_1}+\frac{n_1}{z_1}\right)(x_1-\xi_1) + \frac{n_1}{z_1}(y_1-\eta_1) - \sum_{k=3}^{L}\lambda_k(y_k-y_2) = 0,$$

$$(12)\quad \left(\frac{n_2}{x_2}+\frac{n_2}{z_2}\right)(x_2-\xi_2) + \frac{n_2}{z_2}(y_2-\eta_2) + \sum_{k=3}^{L}\lambda_k(y_k-y_1) = 0,$$

$$(13)\quad \frac{n_1}{z_1}(x_1-\xi_1) + \left(\frac{n_1}{y_1}+\frac{n_1}{z_1}\right)(y_1-\eta_1) + \sum_{k=3}^{L}\lambda_k(x_k-x_2) = 0,$$

$$(14)\quad \frac{n_2}{z_2}(x_2-\xi_2) + \left(\frac{n_2}{y_2}+\frac{n_2}{z_2}\right)(y_2-\eta_2) - \sum_{k=3}^{L}\lambda_k(x_k-x_1) = 0,$$

and for $K = 3, \cdots, L$,

$$(15)\quad \left(\frac{n_k}{x_k}+\frac{n_k}{z_k}\right)(x_k-\xi_k) + \frac{n_k}{z_k}(y_k-\eta_k) + \lambda_k(y_1-y_2) = 0,$$

$$(16)\quad \frac{n_k}{z_k}(x_k-\xi_k) + \left(\frac{n_k}{y_k}+\frac{n_k}{z_k}\right)(y_k-\eta_k) + \lambda_k(x_2-x_1) = 0.$$

For a given K, $K = 3, \cdots, L$, (9), (15), and (16) may be considered as simultaneous linear equations in λ_k, $(x_k-\xi_k)$, $(y_k-\eta_k)$. Solving for λ_k gives $\hat{\lambda}_k$, a function of $(x_1-\xi_1)(x_2-\xi_2)$, $(y_1-\eta_1)$, and $(y_2-\eta_2)$. Substitute this solution in (11) through (14) and obtain four linear simultaneous equations in $(x_1-\xi_1)$, $(x_2-\xi_2)$, $(y_1-\eta_1)$, and $(y_2-\eta_2)$. The solutions, $\hat{\xi}_1$, $\hat{\xi}_2$, $\hat{\eta}_1$, $\hat{\eta}_2$, of these four equations provide BAN estimates of ξ_1, ξ_2, η_1, η_2 from which BAN estimates of α, β, the p's, and the rest of the ξ's and η's are easily obtained.

From (9), (15), and (16) one gets after some simplification

$$(17)\quad \hat{\lambda}_k = \frac{n_k}{D_k}[-(x_1-\xi_1)(y_k-y_2)+(x_2-\xi_2)(y_k-y_1)$$
$$+(y_1-\eta_1)(x_k-x_2)-(y_2-\eta_2)(x_k-x_1)-C_k]$$

where

$$(18)\quad C_k=(y_1-y_2)x_k+(x_2-x_1)y_k-(y_1x_2-y_2x_1)$$

and

$$(19)\quad D_k=x_ky_k(z_2-z_1)^2+x_kz_k(y_2-y_1)^2+y_kz_k(x_2-x_1)^2.$$

Using equation (17) it follows that

$$(20)\quad \sum_{k=3}^{L}\hat{\lambda}_k(y_k-y_2)=-(x_1-\xi_1)\sum_{k=3}^{L}\frac{n_k(y_k-y_2)^2}{D_k}$$
$$+(x_2-\xi_2)\sum_{k=3}^{L}\frac{n_k(y_k-y_1)(y_k-y_2)}{D_k}$$
$$+(y_1-\eta_1)\sum_{k=3}^{L}\frac{n_k(x_k-x_2)(y_k-y_2)}{D_k}$$
$$-(y_2-\eta_2)\sum_{k=3}^{L}\frac{n_k(x_k-x_1)(y_k-y_2)}{D_k}$$
$$-\sum_{k=3}^{L}\frac{n_kC_k(y_k-y_2)}{D_k},$$

with similar expressions for

$$\sum_{k=3}^{L}\hat{\lambda}_k(x_k-x_2),\ \sum_{k=3}^{L}\hat{\lambda}_k(x_k-x_1),\ \text{and}\ \sum_{k=3}^{L}\hat{\lambda}_k(y_k-y_1).$$

Hence substituting in (11), there results

$$(21)\quad (x_1-\xi_1)\left[\left(\frac{n_1}{x_1}+\frac{n_1}{z_1}\right)+\sum_{k=3}^{L}\frac{n_k(y_k-y_2)^2}{D_k}\right]$$
$$+(x_2-\xi_2)\left[-\sum_{k=3}^{L}\frac{n_k(y_k-y_2)(y_k-y_1)}{D_k}\right]$$
$$+(y_1-\eta_1)\left[\frac{n_1}{z_1}-\sum_{k=3}^{L}\frac{n_k(y_k-y_2)(x_k-x_2)}{D_k}\right]$$
$$+(y_2-\eta_2)\left[\sum_{k=3}^{L}\frac{n_k(y_k-y_2)(x_k-x_1)}{D_k}\right]$$
$$=-\sum_{k=3}^{L}\frac{n_k(y_k-y_2)C_k}{D_k},$$

with three more similar equations resulting from the other three substitutions in (12), (13), and (14). Let solutions of these be

$$\hat{\xi}_1 = \hat{\xi}_1(x, y, z),\ \hat{\xi}_2 = \hat{\xi}_2(x, y, z),\ \hat{\eta}_1 = \hat{\eta}_1(x, y, z),\ \hat{\eta}_2 = \hat{\eta}_2(x, y, z).$$

Equations (15) and (16) give,

$$(22) \qquad (x_k - \hat{\xi}_k) = \frac{\hat{\lambda}_k}{n_k}[x_k y_k(z_1 - z_2) + x_k z_k(y_2 - y_1)]$$

$$(23) \qquad (y_k - \hat{\eta}_k) = \frac{-\hat{\lambda}_k}{n_k}[x_k y_k(z_1 - z_2) + x_k z_k(x_2 - x_1)].$$

Hence the estimates $\hat{\xi}_k = \hat{\xi}_k(x, y, z)$, $\hat{\eta}_k = \hat{\eta}_k(x, y, z)$ and $\hat{\zeta}_k = 1 - \xi_k - \hat{\eta}_k$ are found for all $K = 3, \cdots, L$.

It is now apparent that all the estimates may be obtained from the expressions developed above. Thus one gets $\hat{\alpha}$, $\hat{\beta}$, and $\hat{p}_i$; $i = 1, \cdots, L$. The data to be used here are taken from the reported histoplasmin reactions in [1] and are presented in Table 3 together with the estimates obtained. For histoplasmin, $\hat{\alpha}$ was found to be .0811 and $\hat{\beta}$ was equal to .0209. (See Section 7 for a discussion of the tuberculin data in [1].)

TEST OF MODEL

3. As yet nothing has been said concerning a test of Palmer's and Petersen's hypothesis. This is a very important thing and should be done before a judgment of the validity of an hypothesis is made. In this case the χ^2 test is utilized with the previously determined estimates of the ξ's and η's used in obtaining the expected numbers. Call this value of χ^2, χ_a^2. It is given by equation (4) with ξ, η, and ζ replaced by $\hat{\xi}$, $\hat{\eta}$, and $\hat{\zeta}$:

$$(24) \qquad \chi_a^2 = \sum_{i=1}^{L} n_i \left[\frac{(x_i - \hat{\xi}_i)^2}{x_i} + \frac{(y_i - \hat{\eta}_i)^2}{y_i} + \frac{(z_i - \zeta_i)^2}{z_i}\right].$$

In effect what is to be done here is to test the composite hypothesis H_1, which is expressed by equations (1), (2), and (3), when the class of admissible hypotheses consists of all simple hypotheses ascribing positive values to the ξ's, η's, and ζ's subject to the conditions $\xi_i + \eta_i + \zeta_i = 1$, $i = 1, \cdots, L$. Call this latter class the unrestricted set of admissible hypotheses.

The test consists of rejecting H_1 whenever $\chi_a^2 \geq \chi_f^2(\epsilon)$ where $\chi_f^2(\epsilon)$ is the tabled value of χ^2 corresponding to f degrees of freedom and level

TABLE 3

Number of student nurses tested with histoplasmin in specified cities classified as definite, doubtful, and non-reactors, with the corresponding expected numbers as found by various methods

For results of goodness of fit tests, see text and Table 4.

Data from [1], *Public Health Reports*, Vol. 65, No. 1, pp. 1–32.

	CITY	(n_k) NUMBER OF NURSES TESTED	DEFINITE REACTORS ($n_k\xi_k$)				NON-REACTORS ($n_k\eta_k$)				DOUBTFUL REACTORS ($n_k\xi_k$)			
			Obs.	(1) Exp.	(2) Exp.	(3) Exp.	Obs.	(1) Exp.	(2) Exp.	(3) Exp.	Obs.	(1) Exp.	(2) Exp.	(3) Exp.
1	Kansas City, Mo.	1 131	653	657	641.7	594.9	398	402	423.6	446.0	80	72	65.7	90.1
2	Columbus, Ohio	1 330	740	732	748.7	696.1	525	516	504.4	530.7	65	82	76.9	103.2
3	Kansas City, Kans.	532	241	242	241.5	236.6	260	261	259.6	259.9	31	29	30.9	35.5
4	New Orleans	1 290	273	294	279.7	296.0	924	947	920.2	945.3	93	49	90.1	48.7
5	Baltimore	1 971	434	430	432.8	433.5	1 472	1 468	1 473.0	1 471.9	65	73	65.2	65.6
6	Detroit	1 250	158	150	155.0	155.8	1 071	1 064	1 072.3	1 070.8	21	36	22.7	23.4
7	Los Angeles	1 778	189	196	191.0	202.7	1 524	1 531	1 522.8	1 541.8	65	51	64.2	33.5
8	Philadelphia	2 179	228	231	229.2	242.7	1 880	1 885	1 879.2	1 896.8	71	63	70.6	39.5
9	San Francisco	1 111	78	79	78.0	83.9	1 003	1 004	1 003.0	1 012.9	30	28	30.0	14.2
10	Denver	1 329	88	84	86.2	92.5	1 217	1 212	1 218.3	1 221.7	24	33	24.5	14.8
11	Minneapolis	2 419	101	106	101.6	111.1	2 253	2 251	2 252.8	2 287.0	65	56	64.6	20.9
	All cities	16 320	3 183	3 201	3 185.4	3 145.8	12 527	12 547	12 529.2	12 684.8	610	572	605.4	489.4

(1) Expected values as given in [1], page 16.

(2) Expected values, $n_k\xi_k$, $n_k\eta_k$, $n_k\xi_k$, obtained using BAN estimates of unknown parameters.

(3) Expected values, $n_k\xi_k$, $n_k\eta_k$, $n_k\xi_k$, obtained using BAN estimates of unknown parameters, assuming that $\beta = 0$.

of significance ϵ. The number of degrees of freedom is $f = 2L - L - 2$ which, for $L = 11$, is 9.

For the histoplasmin data using the BAN estimates above, $\chi_a^2 = 8.1$ (see Table 4). If the hypothesis H_1 is true, the probability that χ^2 is greater than 8.1 is more than 0.5. This implies that it is not reasonable to reject the hypothesis. In the study which follows, the assumption is made that the hypothesis is true.

The expression (24) used in this test differs from the usual χ^2 in that observed values are used in the denominators instead of expected ones. It is shown in [2] that the test based on this expression is equivalent to the usual χ^2 test in the limit as the number of observations in each sequence (i. e., city) increases.

Some discussion is in order here on the interpretation of the value obtained for χ_a^2. In order to get the estimates and perform the test a simplification of the conditions, $F_k(\xi, \eta) = 0$, on the ξ's and η's was made. The original hypothesis (for which a test is desired) can be written as

$$\eta_k = \xi_k A + B, \text{ for } k = 1, \cdots, L, \tag{25}$$

where A and B are determined from the expressions (1), (2), (3) for $k = 1, 2$ to be

$$A = \frac{\eta_1 - \eta_2}{\xi_1 - \xi_2} \tag{26}$$

$$B = \frac{-(\eta_1\xi_2 - \eta_2\xi_1)}{\xi_1 - \xi_2}. \tag{27}$$

The hypothesis which was actually tested, however, was one specified by the reduced conditions on the ξ's and η's, $F_k^*(\xi, \eta) = 0$. This can be expressed as

$$\eta_k^* = \xi_k^* A^* + B^* + (x_k - \xi_k^*)\left(\frac{\eta_1^* - y_1 - \eta_2^* + y_2}{\xi_1^* - \xi_2^*}\right) + (y_k - \eta_k^*)\left(\frac{\xi_2^* - x_2 - \xi_1^* + x_1}{\xi_1^* - \xi_2^*}\right) \tag{28}$$

where

$$A^* = \frac{\eta_1^* - \eta_2^*}{\xi_1^* - \xi_2^*} \tag{29}$$

and

$$B^* = \frac{-(\eta_1^*\xi_2^* - \eta_2^*\xi_1^*)}{\xi_1^* - \xi_2^*}. \tag{30}$$

The minimum value of χ^2 under the hypothesis as it was desired to be tested was not found exactly but is thought to be between 40 and 50, a value indicating rejection. Hence a contradiction exists between tests

TABLE 4

Estimates of α and β from the histoplasmin data under various hypotheses with corresponding values of χ^2†, and with values of χ_f^2 (.01)§ above which an hypothesis should be rejected at a level of significance of .01

HYPOTHESIS	α	β	χ^2†	DEGREES OF FREEDOM f	$\chi_f^2(.01)$§
1. As given in [1], p. 16 (α, β and the p's are estimated roughly but are in accordance with the original model; χ^2 is not minimized.)	.09	.02	50	9	21.7
2. α, β unspecified (The estimates minimize χ^2 but the original model has been distorted somewhat by using $F_k^*(\xi, \eta) = 0$ instead of $F_k(\xi, \eta) = 0$. Not comparable to 1 above.)	.0811	.0209	8.1	9	21.7
3. $\beta = 0$ (Comparable with 2 above since estimates are obtained using $F_k^*(\xi, \eta) = 0$.)	.1315	0	218.9	10	23.2
4. $\beta = 0$ (The method of maximum likelihood provides the estimates and the original model is retained.)	.1608	0	187.9	10	23.2

† This is χ^2 as computed from equations of form (24).

§ χ^2_f (.01) is such that $P\{\chi^2 \geq, \chi_f^2(.01)\} = .01$ if the hypothesis tested is true; f = number of degrees of freedom.

based on $F_k = 0$ and $F_k^* = 0$. This is to be expected occasionally since the results of the theory tell that only in the limit as sample size increases will these two tests be equivalent (non-contradictory). The acceptance of the hypothesis, H_1, is much more questionable than if

the tests had agreed. Note that the Palmer-Petersen estimates, Table 3, col. (1), were computed for the original hypothesis and gave $\chi^2 = 50.0$.

Out of curiosity the writer took the hypothesis above, $\eta_k = \xi_k A + B$, eliminated A and B using $k = 10$ and 11, instead of $k = 1$ and 2 as done above. This resulted in $F_k^{**} = 0$ which had slightly different numerical coefficients than the $F_k^{*} = 0$. The estimates obtained gave $\hat{\alpha} = .0861$, $\hat{\beta} = .02024$ and provided a χ_a^2 of 10.0. The probability associated with this χ^2 is over 0.30, again indicating non-rejection. This illustrates a weakness of the method of estimation advocated here, namely, the existence of more than one set of estimates. It demonstrates, however, the similarity of the different sets and the good fit obtained from them.

POSSIBILITY OF $\beta = 0$

4. Another related question deals with the supposition that β, the probability of a true negative being a doubtful reactor, is zero. This is somewhat indicated by the histoplasmin data and estimates found therefrom, ($\hat{\alpha} = .0811$, $\hat{\beta} = .0209$). Naturally it is of great importance to examine the question of whether or not the doubtful class contains true negatives.

One may approach this question in a variety of ways: (1) letting $\beta = 0$, one can proceed as in Section 2 obtaining estimates of α and the p's, ξ's, and η's; (2) then a test can be made of this new hypothesis against H_1, which is considered in Section 2; and (3) finally, it is of interest to find the probability of rejecting the hypothesis, $\beta = 0$, when really β equals some small value.

If $\beta = 0$, then, for $i = 1, 2, \cdots, L$, H_2 is written as

$$\xi_i' = p_i'(1 - \alpha'), \tag{31}$$

$$\eta_i' = 1 - p_i' \text{ and} \tag{32}$$

$$\zeta_i' = 1 - \xi_i' - \eta_i'. \tag{33}$$

Maximum likelihood estimates of the parameters can be found easily. They are

$$\tilde{\alpha} = \frac{\sum_{i=1}^{L} n_i z_i}{\sum_{i=1}^{L} n_i (x_i + z_i)} = \frac{\text{number of doubtful reactors}}{\text{number definite} + \text{number doubtful reactors}}, \tag{34}$$

$$\tilde{p} = x_k + z_k, \qquad k = 1, \cdots, L. \tag{35}$$

In order to make the tests mentioned above, however, it is necessary to get BAN estimates in the same way as done in Section 2. Proceeding with this one gets from (31) and (32), for $k = 2, \cdots, L$,

$$F_k'(\xi', \eta') = -\xi_k'\eta_1' + \xi_1'\eta_k' + \xi_k' - \xi_1' = 0, \tag{36}$$

$$\alpha' = \frac{1 - \eta_1' - \xi_1'}{1 - \eta_1'}, \text{ and} \tag{37}$$

$$p_i' = \xi_i'\left(\frac{1 - \eta_1'}{\xi_1'}\right). \tag{38}$$

The reduced form of F_k' is then

$$\begin{aligned} F_k'^*(\xi', \eta') = {} & x_1y_k - y_1x_k + x_k - x_1 + (1 - y_k)(x_1 - \xi_1') \\ & + x_k(y_1 - \eta_1') - (1 - y_1)(x_k - \xi_k') - x_1(y_k - \eta_k') = 0. \end{aligned} \tag{39}$$

In a manner quite analogous to Section 2 the method of Lagrange multipliers is applied and the desired estimates of the parameters are obtained. Call these estimates $\hat{\alpha}'$, $\hat{p}_i'$, $\hat{\xi}_i'$, $\hat{\eta}_i'$, $\hat{\zeta}_i'$; $i = 1, \cdots, L$. In this case $\hat{\alpha}' = 13.15$. See Table 3 for the values of $n_i\hat{\xi}_i'$, $n_i\hat{\eta}_i'$, $n_i\hat{\zeta}_i'$; refer to columns labelled (3).

TEST OF $\beta = 0$

5. A test of the hypothesis that $\beta = 0$ can now be performed using the expression χ_r^2.

$$\chi_r^2 = \sum_{i=1}^{L} n_i\left[\frac{(x_i - \hat{\xi}_i')^2}{x_i} + \frac{(y_i - \hat{\eta}_i')^2}{y_i} + \frac{(z_i - \hat{\zeta}_i')^2}{z_i}\right]. \tag{40}$$

To test H_2 against H_1, one merely computes the difference between the two χ^2 expressions, $\chi_r^2 - \chi_a^2 = \chi_b^2$. This difference tends to be distributed according to the χ^2 distribution with one degree of freedom as the number of observations increases. H_2 is rejected if $\chi_b^2 \geqq \chi_1^2(\epsilon)$, the tabled value of χ^2 with one degree of freedom and level of significance ϵ. Examining the numerical values as shown in Table 4 one sees that if the hypothesis, H_1, is accepted, then for the particular instance, H_2, of H_1, which exists when $\beta = 0$, χ_r^2 is 218.9.

$$\chi_b^2 = 218.9 - 8.1 = 210.8.$$

There is no doubt that this hypothesis, H_2, should be rejected. χ_r^2 was also computed using the maximum likelihood estimates above. $\chi_r^2 = 187.9$, indicating that the approximating done with $F_k'^* = 0$ does

not always result in a χ_r^2 which is minimum with respect to the conditions $F_k' = 0$.

POWER OF TEST OF $\beta = 0$

6. What is the probability of rejecting the hypothesis, $\beta = 0$, (under the general assumptions of this model) when really $\beta = \beta^0$? This probability is called the power of the χ^2 test and depends on α, β, and the p_i's. It can be computed approximately by using the non-central χ^2 distribution as tabled by Evelyn Fix [4]. To use these tables the value of a parameter, λ, must be found. It can be shown [5] that λ is given approximately by

$$\tilde{\lambda} = \chi_r^2 \Big|_{\substack{x=\xi^0 \\ y=\eta^0}} \tag{41}$$

where ξ_i^0 and η_i^0 are the true values of ξ_i and η_i, $i = 1, \cdots, L$.

For the problem at hand the most interesting questions are the following. Suppose that $\sum_{i=1}^{L} n_i = N$, and $Q_i = n_i/N$ is assumed fixed independently of N. Then what is the probability of rejecting H_2 when it is assumed that H_1 as given by (1), (2), and (3) is true? What sample size is required to insure the rejection of H_2, with a probability of, say, .90?

To solve this problem one must first have values for the ξ^0, η^0, ζ^0. These are unknown but under H_1 (assumed true), BAN estimates have been found in Section 2. As an illustration here $\tilde{\lambda}$ has been computed using the $\hat{\xi}$, $\hat{\eta}$, $\hat{\zeta}$ found above as ξ^0, η^0, ζ^0. One has

$$\tilde{\lambda} = N \sum_{i=1}^{L} Q_i \left[\frac{(\hat{\xi}_i^0 - \hat{\xi}_i'(\xi^0, \eta^0, \zeta^0))^2}{\hat{\xi}_i'(\xi^0, \eta^0, \zeta^0)} + \frac{(\hat{\eta}_i^0 - \hat{\eta}_i'(\xi^0, \eta^0, \zeta^0))^2}{\hat{\eta}_i'(\xi^0, \eta^0, \zeta^0)} + \frac{(\hat{\zeta}_i^0 - \hat{\zeta}_i'(\xi^0, \eta^0, \zeta^0))^2}{\hat{\zeta}_i'(\xi^0, \eta^0, \zeta^0)} \right] \tag{42}$$

where $\hat{\xi}_i'(\xi^0, \eta^0, \zeta^0)$, $\hat{\eta}_i'(\xi^0, \eta^0, \zeta^0)$, and $\hat{\zeta}_i'(\xi^0, \eta^0, \zeta^0)$ can be computed as were $\hat{\xi}_i'$, $\hat{\eta}_i'$, and $\hat{\zeta}_i'$, except the originally observed values of x, y, z are replaced by ξ^0, η^0, ζ^0. However, to save labor and recognizing that maximum likelihood estimates are equivalent, for increasing sample size, to the BAN estimates obtained in this paper, the former estimates were used in computing $\tilde{\lambda}$. (Notice that in the first estimation problem, Section 2, maximum likelihood methods result in equations which are very difficult to handle. This accounts for the use of approximate

methods. In the case where $\beta = 0$, maximum likelihood estimates are easily obtainable.) This procedure gives values of $\tilde{\lambda} = 173.9$ and, using $N = 16{,}320$, $\tilde{\lambda}/N = .01066$. As $N \to \infty$, all the $Q_i = n_i/N$ remaining fixed, $\tilde{\lambda}/N$ is independent of N.

From Fix's tables [4], using a level of significance for testing of .05 and 1 degree of freedom, the probability of rejecting the hypothesis $\beta = 0$, when the true values of the unknown parameters are assumed given by the values ξ^0, η^0, ζ^0, is well over .9. If $\tilde{\lambda}/N = .01066$ then a value of $N = 986$ will make $\tilde{\lambda}$ slightly above the 10.509 which is the tabled value for a power equal to .9. For level of significance of .01, $\tilde{\lambda} = 14.879$ corresponds to a power of .9 and N would in this case have to be 1396.

TEST OF $\alpha = \beta$

7. In dealing with the results from tuberculin tests, Palmer and Petersen examine the observed proportions of definite, questionable, and non-reactors, and decide that a fair fit to the data is obtained if α and β are both put equal to .045. Several questions occur when this is considered in connection with the methods of this paper: (i) What would be the outcome of the χ^2 test applied to this hypothesis? (ii) What kind of a fit and what estimates of α and β would result if BAN estimates were found for these parameters? (iii) Is it reasonable to assume α and β are equal but perhaps not equal to .045?

To try to find answers to these questions Table 5 was constructed and is presented here to give the expected values resulting from various hypotheses and various methods of estimation. Included are the estimates published in [1]. Table 6 gives values of χ^2, equation (24), for the various hypotheses and methods of estimation. It is to be noticed that the case where α and β are completely unspecified provides a fit to the data which is much better than that obtained by specifying $\alpha = \beta$. This improvement is obviously more than can be attributed to the decrease in the number of degrees of freedom. (Most of the improvement is due to better agreement between observed and expected in the questionable category.)

Since New Orleans is radically different from the other cities in the percentage of questionable reactors observed, Table 6 was extended eliminating the New Orleans data. The result of this is a markedly better fit in all cases, but α, β unspecified is still the best. If the model now being tested were accepted a priori and the hypothesis $\alpha = \beta$ were to be tested against the class of hypotheses, H_1 (α, β unspecified), then

TABLE 5

Number of student nurses tested with tuberculin in specified cities classified as to definite, questionable, and non-reactors, with the corresponding expected numbers as found by various methods

Data from [1], page 20.

	CITY	(n_k) NUMBER OF NURSES TESTED	DEFINITE REACTORS ($n_k\xi_k$)				
			Obs.	(1) Exp.	(2) Exp.	(3) Exp.	(4) Exp.
1	Kansas City, Mo.	1 131	123	129	125.9	124.1	124.2
2	Columbus, Ohio	1 330	151	144	148.6	149.0	149.1
3	Kansas City, Kans.	532	73	81	75.9	75.5	75.5
4	New Orleans	1 290	171	234	188.6	190.5	190.6
5	Baltimore	1 971	353	365	355.1	357.1	357.4
6	Detroit	1 250	174	173	172.9	173.2	173.4
7	Los Angeles	1 778	333	336	330.7	334.3	334.6
8	Philadelphia	2 179	497	497	484.7	496.6	497.0
9	San Francisco	1 111	207	203	201.9	205.7	205.8
10	Denver	1 329	250	254	182.6	251.4	251.6
11	Minneapolis	2 419	192	177	224.4	189.0	189.1
	All cities	16 320	2 524	2 593	2 491.3	2 546.4	2 548.3

TABLE 5—Continued

	CITY	NON-REACTORS ($n_k\eta_k$)					QUESTIONABLE REACTORS ($n_k\xi_k$)				
		Obs.	(1) Exp.	(2) Exp.	(3) Exp.	(4) Exp.	Obs.	(1) Exp.	(2) Exp.	(3) Exp.	(4) Exp.
1	Kansas City, Mo.	947	951	950.3	955.1	955.9	61	51	54.8	51.8	50.9
2	Columbus, Ohio	1 135	1 126	1 134.0	1 120.0	1 121.0	44	60	47.4	61.0	59.9
3	Kansas City, Kans.	418	427	421.6	432.2	432.5	41	24	34.5	24.3	24.0
4	New Orleans	934	998	965.8	1 040.4	1 041.3	185	58	135.6	59.1	58.1
5	Baltimore	1 506	1 517	1 509.0	1.523.6	1 524.9	112	89	106.9	90.3	88.7
6	Detroit	1 024	1 021	1 023.5	1 019.5	1 020.4	52	56	53.6	57.3	56.2
7	Los Angeles	1 357	1 362	1 356.4	1 362.2	1 363.4	88	80	90.9	81.5	80.0
8	Philadelphia	1 584	1 584	1 583.6	1 582.6	1 584.0	98	98	110.7	99.8	98.0
9	San Francisco	860	858	859.5	854.4	855.2	44	50	49.6	50.9	50.0
10	Denver	1 011	1 015	1 089.4	1 016.7	1 017.6	68	60	57.0	60.9	59.8
11	Minneapolis	2 153	2 133	2 111.3	2 119.2	2 121.0	74	109	83.3	110.8	108.9
	All cities	12 929	12 992	13 004.4	13 025.9	13 037.2	867	735	824.3	747.7	734.5

(1) Expected values as given in [1], page 20.

(2) Expected values obtained using BAN estimates of unknown parameters.

(3) Expected values obtained using BAN estimates of unknown parameters, assuming $\alpha = \beta$.

(4) Expected values obtained using BAN estimates of unknown parameters, assuming $\alpha = \beta = .045$.

TABLE 6

Estimates of α and β from the tuberculin data under various hypotheses with corresponding values of χ^2†, and with values of χ_f^2 (.01)§ above which an hypothesis should be rejected at a level of significance of .01.

HYPOTHESIS	α	β	χ^2†	DEGREES OF FREEDOM f	$\chi_f^2(.01)$§
1. As given in [1], p. 20 ($\alpha = \beta = .045$)	.045	.045	157.1	11	24.7
2. α, β unspecified (The estimates minimize χ^2 but the original model has been distorted by using $F_k^*(\xi, \eta) = 0$ instead of $F_k(\xi, \eta) = 0$.)	.172	.017	55.0	9	21.7
3. $\alpha = \beta$ (In this case $F_k^*(\xi, \eta)$ is the same as $F_k(\xi, \eta)$. The original model is tested with the additional condition that $\alpha = \beta$. Comparable with 2.)	.0458	.0458	141.9	10	23.2
4. $\alpha = \beta = .045$ (Two conditions are imposed; $\alpha = \beta$ and they equal .045. Comparable with 2 and 3 above.)	.045	.045	142.2	11	24.7
2′. α, β unspecified (New Orleans omitted)	.125	.026	25.6	8	20.1
3′. $\alpha = \beta$ (New Orleans omitted)	.043	.043	39.0	9	21.7
4′. $\alpha = \beta = .045$ (New Orleans omitted)	.045	.045	40.5	10	23.2

† This is χ^2 as computed from equations of form (24).

§ $\chi_f^2(.01)$ is such that $P\{\chi^2 \geq \chi_f^2(.01)\} = .01$ if the hypothesis tested is true; f = number of degrees of freedom.

the difference between the two appropriate χ^2's would be distributed in the limit as a χ^2 distribution with one degree of freedom.

$$P\{[\chi^2(\alpha = \beta) - \chi^2(\alpha, \beta \text{ unspecified})] > 39.0 - 25.6\} < .01.$$

The hypothesis, $\alpha = \beta$, should be rejected.

The method of obtaining BAN estimates is the same here as for histoplasmin. The assumption $\alpha = \beta$ is such as to simplify the computations greatly so little labor was involved in this case. The general case, H_1, was of the same difficulty as the histoplasmin one.

SUGGESTIONS FOR MORE GENERAL MODEL

8. As stated in the above section, the χ^2 test rejects the Palmer-Petersen model as regards reactions to the tuberculin test. One is led, therefore, to the consideration of more general models which it is hoped will fit the data more closely and hence provide a better understanding of the situation. A possible extension of the model is based on the fact, as reported in [6], that two people were responsible for making the skin test interpretations in over 95 per cent of the cases. Care was taken that these two interpreters should read the tests uniformly, yet the possibility of systematic differences cannot be ruled out entirely. The model based on such a possibility would state that the proportion of true positive individuals called doubtful by the first reader is $\alpha^{(1)}$ and the proportion of true negatives called doubtful is $\beta^{(1)}$. For the second reader these are $\alpha^{(2)}$ and $\beta^{(2)}$. The problem would now entail estimating $\alpha^{(1)}, \alpha^{(2)}, \beta^{(1)}, \beta^{(2)}, p_1, \cdots, p_2$. Its solution, including the test of the hypothesis, would proceed in a manner analogous to the above.

Palmer and Petersen have advanced as a possible alternative to their model the hypothesis of the existence in some localities, e. g., New Orleans, of a third segment of the population. This segment, called "nonspecific reactors," consists of "persons who have developed a nonspecific sensitivity to the antigen in question." If the distribution of nonspecific reactions according to degree of reaction is distinctly different from those of true negatives and true positives, this might account for the rejection of an hypothesis which did not consider nonspecific reactors.

The above brings forth once more the problems of estimating and testing. It is easily seen that having only the three categories of reactors (definite, questionable, and non-reactors) cannot provide a test since the number of independent parameters is greater than the number of independent observations. In L cities there are $2L$ independent

observations, say, the proportions x_i of definite and y_i of nonreactors, $i = 1, \cdots, L$. But, there are two independent unknown proportions in each city also (the proportions of true positives and true negatives), and in addition, several other parameters. Under these circumstances it is neither possible to estimate all the unknowns nor to make a test of the hypothesis.

In order to obtain estimates and to develop a valid statistical test, one must consider some breakdown of reactors into at least four categories, say, definite ($D+$), essentially definite ($D-$), essentially negative ($N+$) and negative ($N-$) reactors. By so extending the earlier model and proceeding in a manner analogous to Sections 2 and 3, one can estimate the parameters and test the hypothesis involved.

9. Some discussion of estimation procedure is desirable at this point due to the fact that the methods used in this paper have not been seen much since the time Neyman presented his results [2] at the Berkeley Symposium on Mathematical Statistics and Probability in January 1946. In judging any rule of estimation some knowledge of the probability distribution of the estimate is necessary. In a great many instances such probability distributions are not only unknown but practically impossible to find. (In this paper, for example, the estimates, $\hat{\alpha}$ and $\hat{\beta}$, of α and β are functions of the observations and hence vary from sample to sample. The actual distributions of $\hat{\alpha}$ and $\hat{\beta}$ are unknown as are the means and variances.) What is often used in evaluating an estimate is, therefore, not the actual distribution but some other distribution which is equivalent to the true one in some limiting sense. This limiting equivalent distribution is called an asymptotic distribution. Its properties are called the asymptotic properties of the estimate.

One of the main results of Neyman's work [2] is the demonstration that for problems of the type presented in this paper, estimates obtained by minimizing χ^2 directly, by minimizing χ^2 using the reduced conditions, $F_k^*(\xi, \eta) = 0$, *or by maximum likelihood* all have the same asymptotic properties, provided the proportion of observations in each sequence (i. e., city) remains fixed as the total number of observations tends to infinity. All such estimates

(i) are consistent, i. e., tend in probability to the value of the parameter being estimated;

(ii) are asymptotically normal; and

(iii) have minimum asymptotic variance.

Hence, as far as the asymptotic properties are concerned, there is nothing to favor maximum likelihood over minimum χ^2 estimates. This is not to say the actual distributions are the same. They are generally not. Also it is unknown how fast the actual distributions approach the asymptotic ones. It may be possible that some approach faster than others as the number of observations increase. It seems to the author that to get an idea of the comparative properties of different estimation rules in a specific situation there is little that can be done except set up sampling experiments which can approximate the actual distributions empirically. Berkson of the Mayo Clinic has done considerable work on this as related to bioassay.

When it comes to finding estimates in practical situations the author is inclined to use whatever method appears easiest. If maximum likelihood is going to work, it is usually quite simple and the author tries it first. If it appears too complicated, as in the problem of estimating $\alpha, \beta, p_1, \cdots, p_2$, in this paper, then some variation of minimizing χ^2 can be tried. The point seems to be this: since the different estimates are asymptotically equivalent, use the one which is easiest to find. Even the easiest may not be simple, though. The estimates of the parameters and the computation of χ^2 for the hypothesis, α, β unspecified, took about a day to compute. However, to get maximum likelihood estimates by some iteration method seems to the author to be much more difficult and lengthy.

REFERENCES

1. Palmer, Carroll E. and O. Strange Petersen. Studies of pulmonary findings and antigen sensitivity among student nurses, V. *Public Health Reports*, Vol. 65 (Jan. 6, 1950), pp. 1-32.
2. Neyman, J. Contribution to the theory of the χ^2 Test. *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, 1949.
3. ———. Outline of statistical treatment of the problem of diagnosis. *Public Health Reports*, Vol. 62, No. 40 (1947).
4. Fix, Evelyn. Tables of noncentral χ^2. *University of California Publications in Statistics*, Vol. 1, No. 2, pp. 15-19.
5. Taylor, William F. Thesis. (Unpublished).
6. Goddard, J. C., L. B. Edwards, and C. E. Palmer. Studies of pulmonary findings and antigen sensitivity among student nurses, IV. *Public Health Reports*, Vol. 64 (July 1, 1949), pp. 820-846.

ELEVEN MEASURES OF BODY SIZE ON A 1950 SAMPLE OF 15-YEAR-OLD WHITE SCHOOLBOYS AT EUGENE, OREGON*

BY EDWARD O. NEWCOMER AND HOWARD V. MEREDITH

University of Oregon, Eugene, Oregon

ONE of the major objectives of research in physical growth is that of discovering under what conditions childhood growth rates are retarded or accelerated. Weight and stature, and to a lesser extent other measures of body size, have been studied in relation to such variables as birth order, sex, race, secular period, socioeconomic status, geographic location, illness history, diet and physical activity. To date, research workers have accumulated large amounts of data in some instances (*e. g.*, sex differences) but only a paucity of data in others (*e. g.*, influence of physical activity programs on growth).

Few comparative studies have been made "to ascertain whether or not children in various geographic regions of the United States are different with respect to their physical measurements, or to what extent children living in different regions show differences in growth rates" (14, p. 335). This is due mainly to the fact that centers of investigation have been numerous in certain parts of the country and lacking in other parts. Illustrative of the outcome are the extended series of published studies on physical growth available for school children in the northcentral states of Iowa, Illinois, Minnesota, Michigan and Ohio, and the absence of published studies for the northwestern states of Oregon, Idaho, Montana and Wyoming.

The present investigation constitutes an early piece of a long-term program of research initiated in 1950 to study many phases of the physical growth of children residing in Oregon. It deals with eleven measures of body size on a sample of 102 15-year-old white boys drawn from the three junior high schools of Public School District Number 4, Eugene, Oregon.

* From the School of Health and Physical Education, University of Oregon.

AIMS

Specific aims of the investigation were:

1. To study the reliability of anthropometric measurements on junior high school boys. Using composite records from two teams of anthropometrists, reliability analyses were made for stature, weight, stem length, upper limb length, lower limb length, shoulder width, hip width, chest girth, abdomen girth, arm girth and leg girth.

2. To describe 15-year-old Eugene schoolboys of 1950 with reference to the series of eleven selected measurements. For each measurement, the statistical description obtained encompassed both central tendency and variability.

3. To compare 15-year-old, white Eugene boys with white boys living in other parts of the country (*e. g.*, California, Michigan, Maryland) and with boys of other ancestral groups (*e. g.*, American Indian, Mexican, Japanese, Negro). Comparison was made with a total of 17 samples of 15-year-old North American boys all studied since 1930.[1]

SUBJECTS

The subjects were 102 white boys ranging in age from 14 years 9 months to 15 years 3 months. They represented a random sample of the boys of this age enrolled in Public School District Number 4, Eugene, except for the rejection of five individuals exhibiting varying forms of gross body pathology.

All of the subjects were born in North America and practically all were of northwest European ancestry. Forty-six per cent were born in Oregon, 10 per cent in Washington or Idaho, and 5 per cent in California or Nevada. Birthplace for the remaining 39 per cent was scattered, with four boys only born east of the Mississippi River. The mean length of residence in Oregon for the entire sample was 10.3 years.

DEFINITION OF MEASUREMENTS

Stature. Erect body length was measured using an upright board inlaid with millimeter scales and fitted with a wooden square operating in a vertical groove. The subject stood with heels almost together, and with the heels, buttocks, and upper part of the back in contact with the

[1] In geographic and racial comparisons at childhood ages, it is important to utilize studies that are fairly homogeneous for secular period (9, 12).

board. An assistant aided in maintaining this position by placing one hand over the instep and pressing the other against the knees. The arms were permitted to hang at the sides of the body in a natural position. Positioning of the head was in accordance with standard practice, *i. e.*, the head was oriented with the Frankfort Horizontal at right angles to the long axis of the body. As the square was brought down into contact with the vertex, care was taken to see that (a) sufficient pressure was used to crush the subject's hair and (b) the subject kept the plantar surface of his heels firmly upon the floor.

Weight. The subject was requested to stand in the center of the platform on beam-type scales. At the time of the reading, care was taken to observe that the subject's hands were not in contact with the wall, the head portion of the scales, or any other object. The scales were checked frequently and, when necessary, adjusted.

Stem length. The instruments were the same as for stature. The subject sat erect on a horizontal oak bench 40 centimeters in height. His knees were flexed and spread apart, his ankles were crossed, and his hands rested upon his thighs. The posterior aspect of the trunk made contact with the scale at both the sacral and upper thoracic regions. The square was brought down firmly upon the vertex and the measurement recorded as the distance from this point to the surface of the bench.

Upper limb length. This measurement was made with sliding wooden calipers and taken on the left extremity. After extending the extremity at the side of the body and rotating the hand so that its plantar surface faced the lateral side of the left thigh, the distance was found from the lateral point of the acromion process to the most distal point of the middle finger. The instrument was applied from the left rear, care being taken to keep its shaft parallel with the long axis of the limb and to guard against the tendency of many subjects to tilt the shoulder as the limb is extended.

Lower limb length. This measurement was derived by subtraction of each subject's stem length from his stature. The methods of direct measurement of the lower limb length which anthropometrists have used to date, have not proved to be as reliable as the indirect method of taking stature minus stem length.

Shoulder width. Sliding metal calipers having broad flat aluminum branches were used in taking this and the succeeding measurement. The

shoulder breadth obtained was bi-acromial diameter. Each subject stood with the upper extremities extended downward and with the shoulders neither "slumped forward" nor "thrown back." Moderate pressure was applied against the branches of the calipers in order to have the record approximate the skeletal distance between the lateral points of the acromion processes.

Hip width. This measurement was determined as the distance between the lateralmost point of the crest of the right ilium and the corresponding landmark of the left ilium, *i. e.*, as bi-iliocristal diameter. After each branch of the calipers was brought squarely in contact with one of the landmarks, heavy pressure was used in order to approach maximum compression of the overlying soft tissues. In the event that the subject turned his hips as pressure was applied, the measurement was retaken.

Chest girth. The instrument used in taking the four girth measurements was a steel millimeter tape. Chest girth was determined as the circumference of the thorax at the level of the xiphisternal junction. In preparation for measurement each subject stood in a natural manner with head erect and upper extremities relaxed and held slightly away from the sides of the body. The tape was passed around the trunk in a horizontal plane at the xiphisternal level and brought into light contact with the skin. The reading was made as the median girth during normal respiration.

Abdomen girth. Unless the umbilicus projected from the surface contour of the abdomen, this measurement was determined as the circumference of the trunk at the level of the umbilicus. In subjects showing umbilical protrusion, the tape rested against the perimeter of the abdomen at a level immediately above the umbilicus. The posture of the subject, the tension applied to the tape, and the attempt to obtain a midrespiration record were all the same as for chest girth.

Arm girth. The subject stood in a natural manner with the upper extremities hanging in a relaxed condition at the sides of the body and with the left extremity slightly abducted so that the arm was not in contact with the lateral wall of the thorax. The tape was passed around the left arm approximately halfway between the shoulder and the elbow. The area was explored to find the greatest girth at right angles to the long axis of the limb and inferior to any surface indication of the deltoid

muscle. The reading was taken with the tape in sufficiently light contact with the skin to avoid compression of the tissues.

Leg girth. The subject stood with his feet about eight inches apart and his weight distributed equally through both lower limbs. Maximum circumference of the left leg was determined by passing the tape around the leg in the region of the calf and exploring to find the largest reading at right angles to its long axis. As for arm girth, the tape was read at "light contact" tension.

PROCEDURE IN DATA COLLECTION

The subjects were scheduled for measurement individually. Prior to measurement, each boy was briefed regarding the general procedure and the approximate time needed for completion of the examination. Examinations were made in the afternoons between 1:00 and 3:00 o'clock during the months of April, May and June, 1950. All measurements were taken with the subjects wearing nothing but shorts, and at no time was a measuring instrument read where the conditions were considered likely to yield morphologically unsound results. Examples of the latter include chest expansion or breath-holding during the taking of chest girth, contraction of the biceps or triceps muscles during the taking of arm girth, restraining or protruding the abdomen in abdomen girth, and elevating the shoulders in shoulder width.

The data were collected by four anthropometrists conveniently designated as N, P, T, and S.[2] All had been trained in the anthropometric methods currently being used in the child growth research program at the University of Oregon. N and P worked as one team, T and S as a separate team. The procedure may be illustrated as follows: Subject A first went to team N-P, N took the full series of measurements while P acted as assistant and recorder, then P took the series while N acted as assistant and recorder. In instances where N and P did not obtain close agreement, both made additional measurements.[3] The "close agreement" criteria were derived from reliability investigations available for younger boys (3, 8). Specifically, they were agreements within 0.25 kg. for weight, 0.1 cm. for leg girth and hip width, 0.2 cm. for stature and arm girth, and 0.4 cm. for other measurements. It follows that the ultimate values for the N-P team were in some instances the

[2] Edward Newcomer, Everett Peery, Peter Trim, and Paul Sherbina.

[3] Always four measurements were made for chest girth and abdomen girth.

average of four records and in other instances (where it was found unnecessary to take this number) the average of two records.

When examination by team N-P was completed, Subject A passed on to team T-S. This team worked independently of the first team, followed a like procedure, and ultimately obtained a second set of composite values.

RELIABILITY FINDINGS

It is appropriate to commence the analysis of the data with consideration of two methodologic questions. First, do the ultimate values arrived at by the one anthropometric team differ systematically from those arrived at by the other? Is there, for instance, a constant bias in the hip width or arm girth results from the two teams (these biases could exist if the two teams made slightly different laboratory interpretations of "heavy pressure" in taking hip width or "light contact" in taking arm girth)? Secondly, is the magnitude of random variation in the team results reasonably small for all of the dimensions studied? In other words, has chance error in measurement been reduced to the extent that the data for each of the eleven measurements are highly reliable?

Proceeding successively from dimension to dimension, the algebraic difference was secured between every pair of ultimate values from the N-P team and the T-S team. The 102 differences for each dimension next were arranged, with regard to sign, in a frequency distribution. Inspection of these distributions showed no preponderance of differences in one direction over those in the other. For arm girth, 41 differences were positive, 21 zero, and 40 negative; for hip width, 35 differences were positive, 29 zero, and 38 negative. A comparative similarity of positive and negative differences was found for stature, leg girth, stem length, and lengths of the extremities. The distributions which registered the strongest suggestions of bias were those on chest girth and abdomen girth. For chest girth, there were 30 positive differences above 0.2 cm. and 39 negative differences below –0.2 cm.; for abdomen girth, the corresponding numbers were 25 and 34. In neither case did the magnitude of the positive and negative differences indicate marked systematic bias.

Passing to the question of chance variation, this was studied by two methods. The one method consisted of applying the Pearson product-moment method of correlation to the paired values obtained by the two teams, and the other of computing the standard deviation of the dis-

tributions of algebraic differences derived from the paired values. Table 1 gives the findings for each of the eleven traits from both of these methods of studying reliability. It will be seen that:

1. Reliabilities are lowest for shoulder width, abdomen girth, and chest girth. Shoulder width yields the lowest reliability coefficient ($r = .981$) and its standard deviation of team differences ($\sigma_d = 0.32$ cm.) is the largest in relation to variability of the trait. Explicitly, the standard deviation of team differences is equivalent to 15.0 per cent of the standard deviation of shoulder width for this group of 15-year-old boys.

TABLE 1

Anthropometric reliabilities on 15-year-old boys

MEASUREMENT	STANDARD DEVIATION OF DIFFERENCES (σ_d)	RELIABILITY COEFFICIENT (r)
Body weight (kg.)	0.06	.999
Leg girth (cm.)	0.08	.996
Hip width	0.09	.994
Arm girth	0.14	.997
Stature	0.25	.996
Stem length	0.28	.996
Upper limb length	0.29	.992
Lower limb length	0.29	.991
Shoulder width	0.32	.981
Abdomen girth	0.64	.986
Chest girth	0.68	.985

2. Reliability is highest for body weight. The standard deviation of the algebraic differences (0.06 kg.) is equivalent to 0.5 per cent of the standard deviation for weight from this sample of 102 boys. Other reliability sigmas from Table 1 which do not exceed 5.0 per cent of their trait sigmas are those for leg girth, hip width, arm girth, and stature.

FINDINGS ON BODY SIZE OF EUGENE BOYS

The body size values for 15-year-old Eugene boys used in all analyses from here on are the averages of the paired values employed in studying reliability. Stated differently and with particular reference to one dimension, central tendency and variability findings for stature were computed from 102 final values each of which represented a composite

of all the stature records made on a given boy. Similarly for weight, stem length, the four trunk measurements, and the four measurements of the extremities.

Table 2 presents the central tendency and variability results obtained. Findings for central tendency are:

1. Eugene white boys aged 15 years have a mean weight of 60.1 kg. (132.5 pounds) and a mean stature of 169.3 cm. (66.6 inches). Mean stem length or sitting height is 87.6 cm. (34.5 inches).

2. Mean bi-iliocristal hip width is 26.8 cm. (10.5 inches), while biacromial shoulder width is larger by 10.4 cm. (4.1 inches). Trunk circumference means approximate 80 cm. (31.4 inches) at the xiphoid level and 76 cm. (29.9 inches) at the umbilical level.

3. Length of the upper extremity yields a mean of 74.4 cm. (29.3 inches); mean length of the lower extremity (stature minus stem length) is greater by 7.2 cm. (2.8 inches). Similarly, the mean for arm girth is 26.0 cm. (10.2 inches), while the leg girth mean is larger by 8.9 cm. (3.5 inches).

Taken together, the last six values in each row of Table 2 constitute a six-point description of variability which may be regarded as delimiting five normative zones. In recent studies which have employed this procedure for constructing body size norms (10, 11), the successive zones are designated as follows: "Small"—between the minimum value and the 10th percentile; "Moderately small"—between the 10th and 30th percentiles; "Average"—between the 30th and 70th percentiles; "Moderately large"—between the 70th and 90th percentiles; and "Large"—between the 90th percentile and the maximum value. It should be recognized that Table 2 affords no more than tentative norms for 15-year-old Eugene boys; to establish fully adequate zone limits would necessitate a substantially larger sample.

COMPARATIVE FINDINGS

As previously indicated, marked changes are known to have occurred over the last 70 years in the mean body size of North American elementary and secondary school children (9, 12). Since the comparative aim of the present study was oriented in the direction of examining geographic and ethnic differences. it was obligatory to hold the temporal factor as constant as practicable. This was done by aligning the Eugene

TABLE 2

Body size statistics from a sample of 102 Eugene boys 15 years of age

DIMENSION	MEAN	STANDARD ERROR OF MEAN	STANDARD DEVIATION	MINIMUM	PERCENTILES				MAXIMUM
					10	30	70	90	
Body weight (kg.)	60.1	1.1	11.2	33.6	46.7	53.6	65.0	73.9	92.1
Stature (cm.)	169.3	0.8	7.7	148.3	158.3	165.4	173.5	178.9	189.0
Stem length	87.6	0.4	4.2	77.2	81.7	85.9	90.1	92.6	97.5
Shoulder width	37.2	0.2	2.2	30.3	34.3	36.2	38.4	40.3	41.3
Hip width	26.8	0.2	1.8	23.0	24.5	25.8	27.7	29.4	31.9
Chest girth	79.9	0.6	6.1	65.3	72.9	76.2	82.6	89.1	99.0
Abdomen girth	75.9	0.8	7.8	61.8	68.3	71.2	77.7	85.9	105.1
Upper limb length	74.4	0.4	4.1	61.9	69.4	72.5	76.5	79.3	87.7
Lower limb length	81.6	0.5	4.6	70.7	76.2	78.9	83.9	87.0	94.0
Arm girth	26.0	0.3	3.0	18.5	22.5	24.2	27.4	29.8	35.8
Leg girth	34.9	0.3	3.0	27.8	31.3	33.2	36.4	38.5	43.0

means only with means from the literature based upon data collected since 1930.[4]

Seventeen North American samples of 15-year-old non-Oregon boys were assembled for comparative use. These samples included white boys residing in California, Iowa, Maryland and Ontario; boys of Finnish and Italian ancestry in Minnesota; boys of Dutch and Polish ancestry in Michigan; American Chinese, Japanese, and Mexican boys residing in California; American Negro boys in Kansas and California; and American Indian boys in New Mexico and Arizona. Before turning to presentation of the comparative findings, a brief description of each sample will be given.

Stuart and Meredith (17) studied Iowa City white boys attending the University of Iowa laboratory school between 1930 and 1945. These boys were almost entirely of northwest European ancestry and about 60 per cent of them represented the professional and managerial socio-economic groups. Means were obtained for weight, stature, stem length, chest girth, hip width, and leg girth.

Stature and weight means on Los Angeles boys of four ethnic groups measured during 1936-38 were published by Lloyd-Jones (4, 5). The data were gathered at public and private schools as part of a city-wide "School Health Aid Project sponsored by the Los Angeles Board of Education" (5, p. 83). No description of the groups was given other than that carried by the designations, White, Mexican, Negro, and Japanese.

Means for weight, stature, arm girth, and leg girth derived from an extensive 1937-39 survey sponsored by the United States Bureau of Home Economics were reported by O'Brien, Girshick, and Hunt (13). Measurements were made in 16 states and the District of Columbia (none of the states sampled was in the region west of Minnesota and north of California). The subjects were described as "the general run of white American-born boys . . . in public and private schools, on playgrounds, in camps, and in clubs" (p. 2).

Wolff (19, 20) published stature and weight means on white schoolboys of Hagerstown, Maryland. The materials were collected by the United States Public Health Service in annual school surveys made each May "during the years from 1937 though 1940" (19, p. 202).

[4] It would have been preferable to make exclusive use of means from data gathered in the last decade had the volume of available research made this practicable.

Data accumulated through a 1939 stature and weight survey in the elementary schools of Toronto, Canada, were analysed by Keyfitz (2). Measurements were made at public and parochial schools; the subjects were practically all white, with about 85 per cent of northwest European ancestry; there was some selection academically in that 15-year-old boys still attending elementary schools "tend to be backward" (p. 2).

Steggerda and Densen (16) published means for stature and weight on two ethnic groups, Dutch white and Navaho Indian. Their data were amassed between 1931 and 1934 at "schools of Holland, Michigan," and "schools on the Navaho reservation in New Mexico and Arizona" (p. 115). The Dutch white sample included "only those whose ancestors came from the Netherlands" and the Navaho Indian sample only those "who were relatively pure Navaho" (p. 115).

Means for all of the measurements taken on the Eugene boys, with the exception of abdomen girth, were reported by Matheny and Meredith (6) on a group of boys of Finnish ancestry and another group of Italian ancestry. Both groups were measured during 1938-39 at schools within a radius of 50 miles of Hibbing, Minnesota. Although subjects were taken without reference to socioeconomic status, each group was found to be composed of sons "predominantly of skilled or semiskilled workmen" (p. 354).

Data for stature collected in 1939 on boys of Polish descent were analysed by Courtis (1). The data were secured with the assistance of teachers in the schools of Hamtramck, Michigan.

McLendon (7) obtained means for weight, stature, chest girth, hip width, arm girth, and leg girth on Negro boys in attendance 1936-37 at North East Junior High School, Kansas City, Kansas. He described the boys as representing "a wide sampling of the American Negro, being of all degrees of mixture and from all socioeconomic levels" (p. 2).

Means for weight, stature, stem length, shoulder width, and hip width derived from measurement of boys of Japanese lineage were published in 1933 by Suski (18). The data were gathered at "private Japanese day schools in Los Angeles being conducted to supplement the public school education. . . . Both parents of each boy were full-blooded Japanese born in Japan in all instances" (p. 324).

Utilizing data collected in 1935, Wong and Lee (21) reported means for weight, stature, stem length and arm girth on American Chinese boys—offspring of emigrants from the province of Canton. The boys were measured at "public, parochial, and private schools" in Chinatown, San Francisco.

Pitney (15) obtained means for weight, stature, and stem length on Pueblo Indian boys "from 16 pueblos in the Rio Grande valley of New Mexico and 9 Hopi villages in northeastern Arizona" (p. 151). The data were amassed between 1931 and 1934 on all schoolboys "whose

TABLE 3

Stature and weight means for 15-year-old North American schoolboys studied between 1930 and 1950

ETHNIC GROUP	GEOGRAPHIC LOCATION	NUMBER OF SUBJECTS	MEAN STATURE (cm.)	MEAN WEIGHT (kg.)
White*	Eugene, Oregon	102	169.3	60.1
White*	Iowa City, Iowa†	144	167.9	55.6
White	Los Angeles, California	5,202	166.5	55.4
White	Sixteen States	4,740	164.0	52.3
White	Hagerstown, Maryland	850	162.3	51.9‡
White*	Toronto, Canada	1,829	162.3	50.6‡
Dutch	Holland, Michigan	149	168.7	50.5
Finnish§	Hibbing, Minnesota	100	164.4	52.1
Italian§	Hibbing, Minnesota	93	161.4	50.9
Polish	Hamtramck, Michigan	403	158.8	....
Negro	Los Angeles, California	189	166.9	55.4
Negro	Kansas City, Kansas	104	162.6	49.4
Japanese (Suski)	Los Angeles, California	32	161.3	50.0
Japanese	Los Angeles, California	160	160.5	51.1
Chinese	San Francisco, California	80	156.3	45.0
Navaho Indian	New Mexico and Arizona	133	160.5	45.2
Pueblo Indian	New Mexico and Arizona	54	153.5	43.0
Mexican	Los Angeles, California	619	162.2	52.1

* Predominantly boys of northwest European ancestry.

† Private school boys, 60 per cent from professional and managerial classes.

‡ Weighed in indoor clothing but without shoes.

§ Over 80 per cent of fathers were unskilled or semiskilled workmen.

ages were accurately known, whose parents were Pueblo Indians, and who were free from gross physical deformity and defects" (p. 151).

All of the means to which reference has been made are brought together in Tables 3 and 4. Table 3 displays the values for stature and weight, Table 4 those for the remaining measurements. Examination of these tables shows that in no instance are the Eugene schoolboys

TABLE 4

Means (cm.) for eight dimensions of the body on 15-year-old North American schoolboys studied between 1931 and 1950

ETHNIC GROUP	GEOGRAPHIC LOCATION	STEM LENGTH	SHOULDER WIDTH	HIP WIDTH	CHEST GIRTH	LIMB LENGTH		ARM GIRTH	LEG GIRTH
						Upper	Lower		
White*	Eugene, Oregon	87.6	37.2	26.8	79.9	74.4	81.6	26.0	34.9
White*	Iowa City, Iowa‡	86.6	....	26.7	78.3	....	....	....	33.5
White	Sixteen States	....	....	....	....	....	....	24.1†	32.7
Finnish§	Hibbing, Minnesota	84.4	35.5	26.3	74.3	72.3	80.0	24.0	32.4
Italian§	Hibbing, Minnesota	83.4	35.3	26.2	73.3	71.0	78.0	23.6	31.7
Negro	Kansas City, Kansas	....	....	24.5	73.1	....	....	23.6	32.2
Japanese	Los Angeles, California	83.9	33.3	24.3	....	....	....	....	....
Chinese	San Francisco, California	82.8	....	....	....	....	....	20.8	....
Pueblo Indian	New Mexico and Arizona	78.0	....	....	....	....	....	....	....

* Predominantly of northwest European ancestry.

† Measured midway between the acromion and olecranon processes on the right side.

‡ Private school boys, 60 per cent from the professional and managerial class.

§ Over 80 per cent of fathers were unskilled or semiskilled workmen.

surpassed by the schoolboys of other North American studies. Selected findings are as follows:

1. The 15-year-old Eugene boys of the present study are heavier than any of the 16 series of like-age boys with whom they are aligned. Their mean weight is 17.1 kg. (37.7 pounds) greater than that of the Pueblo Indian boys studied by Pitney; 9.2 kg. (20.3 pounds) greater than that of the boys of Italian ancestry studied by Matheny and Meredith; 7.8 kg. (17.2 pounds) greater than that of the white boys from 16 states and the District of Columbia studied by O'Brien, Girshick, and Hunt; 4.7 kg. (10.4 pounds) greater than that of the Los Angeles white and Negro boys studied by Lloyd-Jones; and 4.5 kg. (9.9 pounds) greater than that of the Iowa City laboratory school boys of northwest European ancestry and above average socioeconomic status studied by Stuart and Meredith. All of the differences are statistically significant at the 1 per cent level of confidence.

2. In stature, none of the 17 groups of schoolboys assembled from the literature is taller than the Eugene boys. The means which most closely approximate the Eugene mean are those obtained by Steggerda and Densen on Michigan boys of Dutch ancestry, and by Stuart and Meredith on Iowa private school boys of northwest European ancestry; in neither of these instances are the differences statistically significant for samples of the size used.[5] Other stature comparisons show that the Eugene boys are taller than the Los Angeles white boys from Lloyd-Jones by 2.8 cm. (1.1 inches); taller than the Hagerstown white boys from Wolff and the Toronto white boys from Keyfitz by 7.0 cm. (2.8 inches); taller than the Courtis sample of Hamtramck boys of Polish ancestry by 10.5 cm. (4.1 inches); and taller than the Wong and Lee sample of San Francisco boys of Cantonese lineage by 13.0 (5.1 inches).[6]

3. The Eugene boys have a significantly longer body stem (sitting height) than each group with whom they are aligned in Table 4 except the Iowa laboratory school group. Their mean stem length is higher than that of the Pueblo Indian boys from Pitney by 9.6 cm. (3.8 inches); higher than that of the American Chinese boys from Wong and Lee by 4.8 cm. (1.9 inches); higher than that of boys of Italian descent from

[5] Specifically, $t = 0.6$ for Eugene with Holland, Michigan, and $t = 1.5$ for Eugene with Iowa City, Iowa.

[6] For these comparisons, also the nine more afforded from Table 3 where the Eugene mean exceeds another by over 2.5 cm., the null hypothesis may be rejected at the 1 per cent confidence level.

Matheny and Meredith by 4.2 cm. (1.7 inches); higher than that of boys of Japanese descent from Suski by 3.7 cm. (1.5 inches); and higher than that of boys of Finnish descent from Matheny and Meredith by 3.2 cm. (1.3 inches).

4. For shoulder width, hip width and chest girth, each of the groups placed in juxtaposition with the Eugene boys is significantly smaller with the one exception of the Iowa private school group for hip width. The American Japanese boys from Suski are smaller in shoulder width than the Eugene boys by 3.9 cm. (1.5 inches); the American Negro boys from McLendon are smaller in hip width than the Eugene boys by 2.3 cm. (0.9 inches); and the American white boys from Stuart and Meredith are smaller in chest girth than the Eugene boys by 1.6 cm. (0.6 inches). The means from Minnesota boys of Finnish ancestry are smaller than the Eugene means by 1.7 cm. (0.7 inches) for shoulder width, 0.5 cm. (0.2 inches) for hip width, and 5.6 cm. (2.2 inches) for chest girth.

5. In the four measurements of the extremities, the Eugene boys exceed every group with which they are aligned. The upper extremity is longer for Eugene boys than for Minnesota boys of Finnish and Italian lineage by 2.1 cm. (0.8 inches) and 3.4 cm. (1.3 inches), respectively. Figures from parallel comparisons for length of the upper extremity are 1.6 cm. (0.6 inches) and 3.6 cm. (1.4 inches). Mean arm girth for Eugene boys is larger by 1.9 cm. (0.8 inches) than that for the American white boys studied by O'Brien, Girshick and Hunt; larger by 2.4 cm. (1.0 inches) than that for the American Negro boys studied by McLendon; and larger by 5.2 cm. (2.0 inches) than that for the American Chinese boys studied by Wong and Lee. The mean from the Eugene sample for leg girth is larger than that for the Iowa white boys of northwest European ancestry by 1.4 cm. (0.6 inches); larger than that for the white boys pooled from 16 states and the District of Columbia by 2.2 cm. (0.9 inches); and larger than that for the Minnesota white boys of Finnish ancestry by 2.5 cm. (1.0 inches).

It need not be labored that the discovery of such differences as those enumerated above is a less complex task than that of identifying the specific combinations of variables which have produced the differences. In human biology, phenomena commonly are discovered long before their determinants are understood sufficiently to make possible a full explanation of the phenomena. Nevertheless, inherent in each discovery is some progress toward explanation. From the present vantage point, it can be seen that to explain certain of the obtained differences with reference

to their precise genetic, geographic, temporal, and/or other components will necessitate well-controlled, large scale, long-term patterns of group research. Eugene white boys measured in 1950 are found to be larger than Los Angeles, Hagerstown, and Toronto white boys measured between 1936 and 1940. What variables account for this? How much is due to genetic factors (hereditary differences between the white populations of various localities), how much to geographic differentials (altitude, humidity, solar radiation, medical facilities, food differences), how much to temporal factors (meteorological, migratory, dietary, and/or other changes occurring during the last decade), how much to sampling differences, and so forth? The answer lies ahead—but research is bringing into clearer focus the avenues that will lead toward the answer.

SUMMARY

This study pertains to eleven measures of body size on a sample of 15-year-old Eugene boys. Its three aims are to determine the reliability of anthropometric measurements at this age, to quantitatively describe the body size of the 102 boys studied, and to compare this group of boys with other groups of like age described in the physical growth literature.

Reliability analyses are reported based upon records obtained by two anthropometric teams. Eugene boys are described for central tendency and variability in each of the eleven dimensions. Comparisons are made with 17 other North American studies of fairly recent date.

Acknowledgement is made to Superintendent Clarence Hines, Doctor Madeline Marr and others of the Eugene public schools, who cooperated in making possible the collection of these data, and to the Graduate School of the University of Oregon for a special research grant.

REFERENCES

1. Courtis, S. A. Stature of boys of Polish ancestry. Paper read at Cleveland meeting of Amer. Educ. Res. Assoc., 1939.
2. Keyfitz, N. A Height and Weight Survey of Toronto Elementary School Children, 1939. *Ottawa*: Department of Trade and Commerce, Dominion Bureau of Statistics, 36 pp., 1942.
3. Knott, Virginia B. Physical Measurement of Young Children: A Study of Anthropometric Reliabilities for Young Children Three to Six Years of Age. *Univ. Iowa Stud. Child Welf.*, 18, 3: 99 pp., 1941.
4. Lloyd-Jones, Orren. California tall children. *Amer. J. Dis. Child.*, 60: 11-21, 1940.

5. ———. Race and stature: a study of Los Angeles school children. *Res. Quart.*, 12: 83-97, 1941.
6. MATHENY, WALTER D., and HOWARD V. MEREDITH. Mean body size of Minnesota schoolboys of Finnish and Italian ancestry. *Amer. J. Phys. Anthrop.*, 5: 343-355, 1947.
7. MCLENDON, JOHN B., JR. Standards of nutritional status for Negro high school boys. Unpublished M. A. thesis, Univ. Iowa, 1937.
8. MEREDITH, HOWARD V. The reliability of anthropometric measurements taken on eight- and nine-year-old White males. *Child Develop.*, 7: 262-272, 1936. (See also, MARSHALL, EVERETT L. The objectivity of anthropometric reliabilities taken on eight- and nine-year-old White males. *Child Develop.*, 8: 249-256, 1937.)
9. ———. The stature and weight of United States children as influenced by racial, regional, socioeconomic and secular factors. *Amer. J. Dis. Chld.*, 62: 909-932, 1941.
10. ———. A "Physical Growth Record" for use in elementary and high schools. *Amer. J. Pub. Health*, 39: 878-885, 1949.
11. ———. Body size norms for children four to eight years of age. *J. Pediat.*, 37: 183-189, 1950.
12. ——— and E. MATILDA MEREDITH. The stature of Toronto children half a century ago and today. *Hum. Biol.*, 16: 126-131, 1944.
13. O'BRIEN, RUTH, M. A. GIRSHICK, and E. P. HUNT. Body Measurements of American Boys and Girls for Garment and Pattern Construction. *Washington*: Bureau of Home Economics, United States Department of Agriculture, *Miscellaneous Publication* No. 366, 141 pp., 1941.
14. PALMER, CARROLL E., and SELWYN D. COLLINS. Variations in physique and growth of children in different geographic regions of the United States. *Pub. Health Rep.*, 50: 335-347, 1935.
15. PITNEY, ELIZABETH H. The physical size and growth of Pueblo Indian children. Unpublished Ph. D. thesis, Yale Univ., 1940.
16. STEGGERDA, MORRIS, and PAUL DENSEN. Height, weight, and age tables for homogeneous groups, with particular reference to Navaho Indians and Dutch White. *Child Develop.*, 7: 115-120, 1936.
17. STUART, HAROLD C., and HOWARD V. MEREDITH. Use of body measurements in the school health program: Part II. Methods to be followed in taking and interpreting measurements and norms to be used. *Amer. J. Pub. Health*, 36: 1373-1386, 1946. (Also further unpublished findings.)
18. SUSKI, P. M. The body build of American-born Japanese children. *Biometrika*, 25, Parts 3 and 4: 323-352, 1933.
19. WOLFF, GEORG. Further results on the trend of weight in White school children. *Child Develop.*, 12: 183-205, 1941.
20. ———. A study of height in White school children from 1937 to 1940 and a comparison of different height-weight indices. *Child Develop.*, 13: 65-77, 1942.
21. WONG, THOMAS A., and SAMUEL D. LEE. Anthropometric Study of the San Francisco Chinese. Final Report, Official Statistical Project No. 65-3-4229, WPA Work Project No. 3330, 25 pp., 1936.

THE SEX RATIO OF STILLBIRTHS RELATED TO CAUSE AND DURATION OF GESTATION*

AN INVESTIGATION OF 7,066 STILLBIRTHS

BY THOMAS McKEOWN AND C. R. LOWE

The Department of Social Medicine,

University of Birmingham

ALTHOUGH no direct evidence of the sex ratio at conception is available, estimates have frequently been made from the observed ratios of live births, stillbirths, total births and abortions.[1] The considerable variation in the estimates is mainly due to uncertainty about abortions, for the sex ratios of live births, stillbirths and total births are well known. Reliable information about all foetuses lost from the uterus might be expected to establish the ratio at conception; but while the sex ratio of stillbirths is not in doubt, it is unfortunately very difficult to obtain accurate data on abortions, particularly in the early months of gestation for which existing records are most deficient. This difficulty led us to enquire whether more could be learned from examination of stillbirths, which in Great Britain are distinguished from abortions by the time (28 weeks) fixed arbitrarily for their notification.

A further reason for reconsidering the sex ratio of stillbirths was suggested by an investigation of the variation of the sex ratio with maternal age, which drew attention to the importance of individual causes of stillbirth and abortion (Lowe and McKeown, 1950). Recognizing that the sex ratio of stillbirths explains the difference between ratios of live and total births (as the sex ratio of abortions, if known, would explain any difference between the ratio of total births and the

* This research was assisted by a grant from the Birmingham University Students' Social Services Fund.

[1] In conformity with practice in Great Britain, we describe as abortions all foetuses born dead before the 28th week of gestation, and as stillbirths all foetuses born dead after the 28th week of gestation. Stillbirths are notified; abortions are not.

conception ratio), we examined the sex ratios of different causes of stillbirth. It was observed (a) that different causes of stillbirth have different sex ratios, and (b) that the proportionate contribution of each cause varies with maternal age. Although stillbirths (all causes) exhibited a fairly consistent age trend, there was no evidence of age variation of the sex ratio of any single cause which could be examined in substantial numbers. This suggested that variations in the sex ratio of stillbirths with maternal age, or indeed with any other variable, might be explained wholly or in part by changes in the composition of stillbirths. And since the distinction between stillbirths and abortions depends merely on the date of notification, the same considerations may be expected to apply to abortions.

From the point of view of the sex ratio at conception, the most important issue is the change in the sex ratio of stillbirths and abortions with duration of gestation. In this communication we examine the sex ratio of stillbirths by cause and duration of gestation, and discuss the bearing of the data on estimates of the sex ratio at conception.

SOURCES OF DATA

For the years 1936-1949 we obtained from city records data for all stillbirths notified in Birmingham, England.[2] During this period the Maternity and Child Welfare Department, assisted by the attendant at the birth (doctor or midwife), completed a record card in respect of each stillbirth; we are here concerned with only two of the items entered, the cause of stillbirth and the estimated duration of gestation. For the first seven years (1936-1942) duration of gestation is given in months, and for the remaining seven years (1943-1949) in weeks. In both cases the estimate is based on the date of the last menstrual period, and is of course subject to the errors inseparable from this method.

In the 14-year period there were 7,319 stillbirths of which 253, from multiple pregnancies, have been excluded. The analysis which follows is based on the 7,066 stillbirths from single pregnancies.

[2] National statistics for Scotland give the cause of stillbirths, but not the duration of gestation. National statistics for England and Wales give neither cause of stillbirths nor duration of gestation.

SEX RATIO OF STILLBIRTHS RELATED TO CAUSE AND DURATION OF GESTATION

The 7,066 stillbirths have been grouped by cause in four classes (see Appendix):

(1) disease in or accident to the mother;

(2) foetal malformation;

(3) difficult labour;

(4) ill-defined or unknown cause.

TABLE 1

Sex ratio of stillbirths related to cause and duration of gestation, Birmingham, 1936-1949

STILLBIRTH ATTRIBUTED TO:	DURATION OF GESTATION				
	7 months (28-32 wks.)	8 months (33-37 wks.)	9 months and over (38 wks. and over)	Unspecified	Totals
Disease in or accident to the mother	55.9 (515)	54.0 (631)	56.2 (660)	53.6 (97)	55.3±1.1 (1903)
Foetal malformation	29.4 (252)	36.2 (356)	42.9 (469)	42.1 (38)	37.7±1.5 (1115)
Difficult labour	57.1 (77)	62.4 (173)	62.4 (1761)	61.8 (110)	62.2±1.1 (2121)
Ill-defined or unknown cause	53.4 (395)	55.6 (399)	55.0 (1003)	56.2 (130)	54.9±1.1 (1927)
Totals	49.8±1.4 (1239)	51.3±1.3 (1559)	57.1±0.8 (3893)	55.7±2.6 (375)	54.5±0.6 (7066)

This corresponds with the classification used in the Annual Reports of the Registrar-General for Scotland, although the first three classes are more homogeneous than those of the Registrar-General, and consequently the fourth class ("ill-defined or unknown cause") is larger.

Table 1 gives sex ratios[3] by cause of stillbirth and duration of gestation (see also Fig. 1). We may note:

[3] Sex ratio is expressed as the percentage of males $[M/(M+F)] \times 100$.

(a) The sex ratio of stillbirths (all causes) increases as the duration of gestation increases (from 49.8 at seven months to 57.1 at nine months and over);

(b) Sex ratios of the four classes of stillbirth differ from one another (the differences are exhibited in each month of gestation as well as in the totals);

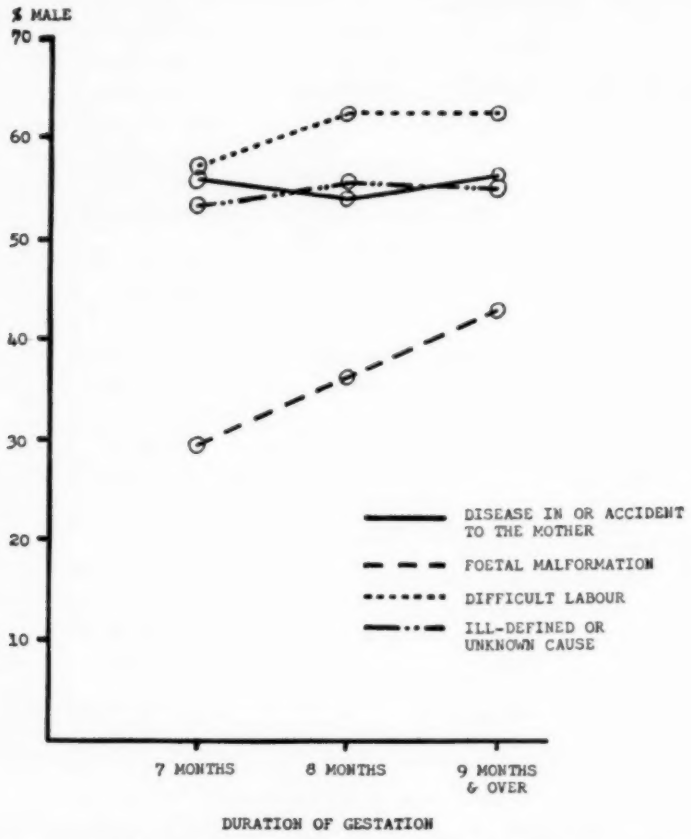

FIG. 1. SEX RATIO OF STILLBIRTHS RELATED TO CAUSE AND DURATION OF GESTATION, BIRMINGHAM, 1936-1949

(c) Only one of the four classes (foetal malformation) shows a consistent relationship between sex ratio and month of birth, but in this case the increase with duration of gestation is considerable (from 29.4 at seven months to 42.9 at nine months and over).

Let us next enquire whether the change in the sex ratio of stillbirths (all causes) with duration of gestation is in part explained by changes in the composition of stillbirths, since, as noted above, different causes of stillbirth have different sex ratios. Table 2 gives the percentage distribution by cause of stillbirth for each month of gestation (7, 8, and 9 and over). As would be expected, the proportion of stillbirths

TABLE 2

Percentage distributions of stillbirths by cause at different durations of gestation, Birmingham, 1936-49

STILLBIRTH ATTRIBUTED TO:	DURATION OF GESTATION (months)				
	7	8	9 and over	Unspecified	Totals
Disease in or accident to the mother	41.6	40.5	17.0	25.9	26.9 (1903)
Foetal malformation	20.3	22.8	12.0	10.1	15.8 (1115)
Difficult labour	6.2	11.1	45.2	29.3	30.0 (2121)
Ill-defined or unknown cause	31.9	25.6	25.8	34.7	27.3 (1927)
Totals	100.0 (1239)	100.0 (1559)	100.0 (3893)	100.0 (375)	100.0 (7066)

attributed to each cause changes with the duration of gestation; for example "difficult labour" accounts for 6.2 per cent of deaths (77 out of 1239) in the seventh month, and for 45.2 per cent of deaths (1761 out of 3893) at nine months and over.

This suggests that the increase in the sex ratio of stillbirths (all causes) between the seventh and ninth months reflects:

(i) the different sex ratios of the four classes of stillbirth;

(ii) the change with duration of gestation in the proportion of stillbirths attributed to each class; and

(iii) the change with duration of gestation of the sex ratio of the class "foetal malformation."

We can readily eliminate the effect of (i) and (ii) by standardizing

the sex ratios for changing distribution of causes. Sex ratios have been calculated for stillbirths (all causes) in each month of gestation by applying the sex ratios for each cause in that month to the proportions of the four classes (26.9, 15.8, 30.0, 27.3) in all stillbirths (see Table 2) The adjusted sex ratios [4] are then:

7 months	51.3	(49.8)
8 months	54.1	(51.3)
9 months and over	55.6	(57.1)

The increase which remains after this adjustment is mainly accounted for by the change with duration of gestation in the sex ratio of the class "foetal malformation." Since this class is composed of different malformations whose sex ratios differ from one another, it is of interest to examine the sex ratios of the individual malformations in relation to duration of gestation.

SEX RATIO OF FOETAL MALFORMATIONS RELATED TO CAUSE AND DURATION OF GESTATION

The class "foetal malformation" contains six sub-groups (see Appendix):

(a) anencephalus alone;
(b) anencephalus with spina bifida;
(c) hydrocephalus alone;
(d) hydrocephalus with spina bifida;
(e) spina bifida alone;
(f) other malformations.

Numbers do not permit a separate examination of the sex ratio of each sub-group by duration of gestation, and we have therefore combined (a) and (b) as "anencephalus," (d) and (e) as "spina bifida," and (c) and (f) as "other foetal malformations." Table 3 (see also Fig. 2) shows that the sex ratios of "anencephalus" and "spina bifida" increase with duration of gestation; the sex ratio of "other foetal malformations" exhibits a slight but insignificant decrease.

We cannot, however, entirely attribute the change in the sex ratio of the class "foetal malformation" to changes in the ratios of the

[4] Numbers in brackets are the unstandardized sex ratios.

individual malformations, without considering once again the possible effect of alterations in their relative proportions. For example, Table 3 shows that " other foetal malformations " accounts for 57 (22.6 per cent) of 252 malformations in the seventh month, and for 205 (43.7 per cent) of 469 malformations at nine months and over. Standardizing as before

TABLE 3

Sex ratio of stillbirths attributed to foetal malformations related to duration of gestation, Birmingham, 1936-49

STILLBIRTH ATTRIBUTED TO:	DURATION OF GESTATION (months)				
	7	8	9 and over	Unspecified	Totals
Anencephalus*	22.1 (172)	31.3 (230)	38.6 (145)	21.4 (14)	30.1±1.9 (561)
Spina bifida†	21.7 (23)	28.6 (42)	36.1 (119)	50.0 (8)	33.3±4.0 (192)
Other foetal malformations	54.4 (57)	53.6 (84)	49.8 (205)	56.3 (16)	51.7±2.6 (362)
Totals	29.4±2.9 (252)	36.2±2.5 (356)	42.9±2.3 (469)	42.1±8.0 (38)	37.7±1.5 (1115)

* Includes anencephalus with spina bifida.

† Includes spina bifida with hydrocephalus.

for changing distribution of causes we get sex ratios [4] for the three periods as follows:

7 months	32.5	(29.4)
8 months	38.1	(36.2)
9 months and over	41.8	(42.9)

On this evidence we may conclude that the increase in the sex ratio of the class " foetal malformation " with duration of gestation is mainly due to the change in the sex ratios of the individual malformations of which it is composed, in particular of anencephalus and spina bifida.

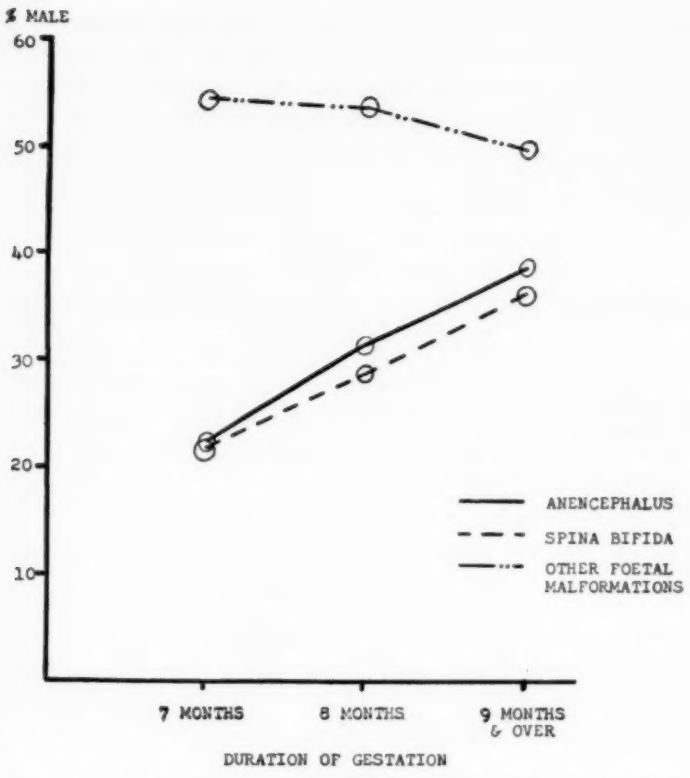

FIG. 2. SEX RATIO OF STILLBIRTHS ATTRIBUTED TO FOETAL MALFORMATIONS RELATED TO DURATION OF GESTATION, BIRMINGHAM, 1936-1949

TABLE 4

*Sex ratios of live born and stillborn malformations**

MALFORMATION	PERCENTAGE MALE		
	Live Births (*a*)	Stillbirths (*b*)	Difference (*a*) — (*b*)
Anencephalus	52.6	30.8	21.8±10.7
Spina bifida	46.1	36.1	10.0± 5.5

* Taken from Record and McKeown (1949).

It is of interest to note (Table 4) that for both these malformations the sex ratio is higher for live births than for stillbirths (Record and McKeown, 1949). This observation is quite consistent with the foregoing, since on the average liveborn malformations are born later in gestation than stillbirths.

Sex ratios of stillbirths lost from the uterus during each month of gestation are of course influenced by the relative proportions of males and females in the uterus. Unfortunately, for most malformations we are unable to estimate numbers in the uterus, because the considerable number of affected liveborn individuals is unknown. But the proportion of anencephalics which survives birth is so small that in this case we can without serious error compute from the stillbirths the number of anencephalic individuals retained in and lost from the uterus. This information is provided in Table 5, which uses records for the years 1943-1949 for which duration of gestation can be given in fortnights. Perhaps the simplest interpretation of this evidence is as follows:

(a) There are more anencephalic females than males in the uterus at 28 weeks (sex ratio 26.4);

(b) Between 28 and 30 weeks the loss of females is *relatively* greater than the loss of males; and

(c) In consequence of (b) the sex ratio of anencephalic foetuses in the uterus is higher at 30 weeks (26.9) than at 28 weeks.

The same explanation can be given for the regular increase in the sex ratio of anencephalics retained in the uterus at each subsequent fortnight.

SEX RATIO OF MACERATED STILLBIRTHS

We have so far considered the sex ratio of stillbirths by cause, without reference to the condition of the foetus. In Table 6 sex ratios of macerated and not macerated stillbirths are given by duration of gestation and cause of stillbirth. There were no macerated foetuses among stillbirths attributed to difficult labour (as might be expected) and therefore this cause of death does not appear. For every other cause of stillbirth the sex ratio of the totals is higher for macerated than for not macerated stillbirths. Numbers do not justify detailed comment on sex ratios by cause of death, though we may note that with one exception ("other foetal malformations") sex ratios of macerated stillbirths increase with duration of gestation. (For "anencephalus and

TABLE 5

Sex ratio of anencephalics retained in and lost from the uterus at different periods of gestation, Birmingham, 1943–49

DURATION OF GESTATION (weeks)	ANENCEPHALICS IN THE UTERUS AT BEGINNING OF EACH PERIOD*			ANENCEPHALICS LOST FROM THE UTERUS DURING EACH PERIOD				
				Number lost		Percentage lost		Sex ratio
	Male	Female	Sex ratio	Male	Female	Male	Female	
28 —	73	203	26.4	3	13	4.1	6.4	18.7
30 —	70	190	26.9	9	32	12.9	16.8	22.0
32 —	61	158	27.9	7	31	11.5	19.6	18.4
34 —	54	127	29.8	10	36	18.5	28.3	21.7
36 —	44	91	32.6	19	45	43.2	49.5	29.7
38 —	25	46	35.2	10	20	40.0	43.5	33.3
40 —	15	26	36.6	15	26	100.0	100.0	36.6

* Excludes a small number of live born anencephalics and six stillborn anencephalics for which the duration of gestation was unknown.

Sex ratios of macerated and not macerated foetuses related to duration of gestation, Birmingham, 1936–49

DURATION OF GESTATION (months)

Sex ratios of macerated and not macerated foetuses related to duration of gestation, Birmingham, 1936–49

STILLBIRTH ATTRIBUTED TO:	CONDITION OF STILLBIRTH	DURATION OF GESTATION (months) 7	8	9 and over	Unspecified	Totals
Disease in or accident to the mother	Macerated	55.4 (233)	60.4 (288)	63.5 (277)	52.6 (38)	59.7 (836)
	Not macerated	56.4 (282)	48.7 (343)	50.9 (383)	54.2 (59)	51.8 (1067)
Foetal malformation						
(a) Anencephalus and spina bifida	Macerated	22.0 (50)	33.8 (68)	48.1 (54)	75.0 (4)	35.8 (176)
	Not macerated	22.1 (145)	29.9 (204)	34.8 (210)	22.2 (18)	29.5 (577)
(b) Other foetal malformations	Macerated	55.3 (38)	56.8 (44)	52.8 (106)	100.0 (6)	55.7 (194)
	Not macerated	52.6 (19)	50.0 (40)	46.5 (99)	30.0 (10)	47.0 (168)
Ill-defined or unknown cause	Macerated	52.6 (287)	55.7 (298)	58.3 (618)	54.0 (50)	56.2 (1253)
	Not macerated	55.6 (108)	55.4 (101)	49.9 (385)	57.5 (80)	52.5 (674)
Totals	Macerated	51.3 (608)	55.6 (698)	58.6 (1055)	57.5 (98)	55.9 (2459)
	Not macerated	47.1 (554)	44.2 (688)	47.0 (1077)	50.9 (167)	46.5 (2486)

spina bifida" the increase is considerable, though in this case the sex ratio of not macerated foetuses exhibits the same trend.) When all causes of stillbirth are combined, the increase in the sex ratio of macerated foetuses with duration of gestation is quite definite; the sex ratio of not macerated foetuses exhibits no regular trend.

These observations, for which we offer no explanation, presumably mean that after death males are more likely to be retained in the uterus than females.

DISCUSSION

As already indicated, since ratios of stillbirths, live births and total births are not in doubt, the difficulty in establishing an acceptable estimate of the sex ratio at conception derives from uncertainty about the incidence and sex ratio of abortions earlier than the 28th week of gestation. We may therefore profitably enquire whether the information elicited for stillbirths has any bearing on the probable sex incidence of abortions, and whether it enables us to select with any confidence from the wide range of values which the literature offers for the conception ratio.

Estimates of the sex ratios of abortions and stillbirths by month of gestation have been assembled from the literature in Table 7. It will be best if we consider stillbirths separately, since they are more reliably recorded than abortions. On two points there is surprisingly good agreement:

(1) The sex ratio at the ninth month is about 57;

(2) The sex ratio decreases between the ninth and seventh months. A third point on which there is less general agreement, is suggested by some of the more recent investigations (Bayer, 1939; Tietze, 1948) including our own;

(3) At the seventh month the sex ratio does not differ significantly from 50.

In support of (3) it may be noted that in Great Britain the sex ratio at the seventh month cannot be much above 50 (if we accept (1) and (2)) since (a) the sex ratio of all stillbirths is firmly established at or just below 54,[5] and (b) there are fewer stillbirths in the seventh than in the eighth or ninth months (Table 1). Variation can of course be

[5] England and Wales, 53.82 (1939-47); Scotland, 54.00 (1939-46).

TABLE 7

Sex ratios of abortions and stillbirths by month of gestation

SOURCE OF DATA AND REFERENCE	NUMBER OF CASES	MONTH OF GESTATION						
		3	4	5	6	7	8	9
Birmingham Stillbirth Register, 1936–1949 This communication	6,691	....	....	....	....	49.8	51.3	57.1
Notifications in Budapest, 1903–5 Auerbach (1912)	4,067	....	69.6	62.0	53.7	53.6	....	....
Hospital Obstetrical Practice, Heidelberg Bayer (1939)	2,309	91.9	68.9	51.4	52.5	49.2	53.6	57.3
Quoted by Parkes (1926) Bertillon (1893)	.....	....	64.3	54.3	52.8	53.8	51.6	56.8
Notifications in a selected area of U. S., 1925–34 Ciocco (1938)	163,447	78.3	66.8	58.3	55.1	52.7	55.5	57.4
Notifications in Paris (1891–1902)	59,688	....	63.8	54.8	53.9	54.3	55.2	57.3
Notifications in District of Columbia (1874–1902)								
White	4,400	....	59.2	59.7	52.6	53.3	51.5	56.3
Coloured	7,500	....	55.0	60.3	55.4	55.0	52.8	55.9
Nicholls (1907)								
Carnegie Collection, Washington	1,410	54.8	54.0	52.3	46.7	52.0	57.1	62.6
	507	....	45.6	51.3	53.8	55.3	....	....
	5,787	48.3	51.9	53.6	52.3	50.2	....	....
Schultz (1921); Streeter (1920); Tietze (1948)								
Notifications in Vienna, 1893–191	35,340	....	73.9	60.5	53.5	54.4	....	....
Notifications in Budapest, 1903–1907	6,731	....	69.8	60.6	54.4	53.9	....	....
Notifications in Paris, 1901–1909 Tschuprow (1915)	42,853	....	66.7	56.6	55.1	54.7	....	....

expected between different countries according to the composition of stillbirths; for example, the ratio given by Ciocco (1938) for the seventh month (52.9) may have been valid for a selected area of the United States in the period 1925-1934, though it is not consistent with recent ratios for stillbirths in Scotland and in England and Wales.

It is with sex ratios earlier than the seventh month that serious difficulty appears, as is evident from the wide variation in figures given for abortions. Most estimates are, however, over 50, and with few exceptions (Boldrini, 1930) authors have been concerned with the extent of the male preponderance, without doubting its existence. Before venturing an opinion on this matter we should consider carefully two difficulties common to all investigations of the sex ratio of human abortions.

The first is the difficulty of obtaining a representative selection of abortions large enough to give reliable ratios. There can be little doubt that in most series the proportion of abortions reported in the early months of pregnancy is very low, either because the sex is not recorded, or more commonly because the aborted material is never seen by a doctor. Ciocco's data for selected states show that the proportion of notified abortions in which sex was recognized was only 28 per cent in the second month, and was below 90 per cent until the fifth month. Museum specimens are in a special category, but there must always be some doubt about their randomness, even in the case of the important Carnegie collection which has been examined several times (Streeter, 1920; Schultz, 1921; Tietze, 1948).

The low proportion of abortions recorded would of course be less serious if there were no particular reason for doubting its randomness. But a very important reason for doubt is suggested by consideration of the limited information available on the pathology of abortions. The pathologist interested in cause of death faces of course the same difficulty in collecting representative material as the pathologist interested in the sex, but he requires for his purpose much smaller numbers. The most useful data are provided by Hertig (1943), who investigated the pathology of 1,416 consecutive spontaneous abortions submitted by hospitals or physicians in the Boston area. His classification of 1000 of the cases (selected for completeness of clinical history and pathological material) is as follows:

1. *Ovular factors*		
(i) Pathologic ova, with absent or defective embryos	489	
(ii) Embryos with localized anomalies	32	
(iii) Placental abnormalities	96	
	——	617
2. *Maternal factors*		
(i) Criminal abortions	21	
(ii) Uterine abnormalities	64	
(iii) Febrile and inflammatory diseases	20	
(iv) Miscellaneous	12	
(v) Anatomically normal ova (classified)	265	
(vi) Trauma (automobile accident)	1	
	——	383
		1000

Hertig's examination of spontaneous abortions suggests that about half the embryos (48.9 per cent) are either absent or so defective that there can be no question of sex-identification. Moreover it must be remembered that a considerable proportion of abortions (half is a usual estimate) are induced, and are in consequence rarely submitted for examination unless clinical complications bring them to hospital. If we accept Hertig's evidence it is clear that the abortions upon which sex is reported are highly selected, and may be drawn from not more than one-quarter of the cases. This is particularly serious because this quarter very largely excludes pathological ova, in which causes of death may be highly sex specific.

The second difficulty is in determining the sex of young embryos. Schultz (1921) drew attention to the difficulty of early sex differentiation from the external genitalia; Spaulding (1921) and Wilson (1926) suggested that before 11 weeks even the opinion of experts is unreliable when based on macroscopic examination, and mistakes may easily be made until the fifth month. This question has recently been reviewed by Tietze (1948), who considers we can have little confidence in existing estimates when we recall that in most cases sex has been determined by persons who have had no specialized knowledge of embryology.

These observations are all of course more pertinent in the early months, and it may be argued that they hardly justify rejection of published estimates for the fifth and sixth months, which are almost without exception above 50 (Table 7). But although the precise incidence of abortions is unknown, most writers are agreed that it is very

much higher in the early than in the late months of pregnancy, and that at least half of all abortions occur in the first three months (Nichols, 1907; Auerbach, 1912; Carvallo, 1912; Brunner and Newton, 1939). Even if it is true, therefore, that the sex ratio is above 50 in the fifth and sixth months, the gross inadequacy of information about the earlier months must leave the sex ratio of all abortions, and hence the sex ratio at conception, an open question.

TABLE 8

Estimated sex ratio of abortions assuming that equal numbers of males and females are conceived

		MALES	FEMALES	TOTAL	SEX RATIO
Number of foetuses at conception		50 000	50 000	100 000	50.0
Abortion rates:					
15%	Abortions	6 225	8 775	15 000	41.5
	Total births*	43 775	41 225	85 000	51.5
25%	Abortions	11 375	13 625	25 000	45.5
	Total births*	38 625	36 375	75 000	51.5
35%	Abortions	16 525	18 475	35 000	47.2
	Total births*	33 475	31 525	65 000	51.5

* The sex ratio of total births (51.5) is the observed ratio for England and Wales (1939-47).

In conclusion, since the sex ratio of total births is known, it is of interest to enquire what the sex ratio of abortions must be if equal numbers of males and females are conceived. For this purpose we require to know the proportion of pregnancies terminated by abortion, and estimates in the literature are between 12 per cent (Wiehl, 1938) and 44 per cent (Brunner, 1941). The sex ratio of total births in England and Wales (1939-1947) was 51.5, and assuming that 15%, 25%, or 35% of pregnancies end in abortion, we get values of 41.5, 45.5, or 47.2 respectively for the sex ratios of abortions (Table 8). Except in so far as our knowledge of the mechanism of sex determination disposes us to expect a conception ratio of 1 : 1, we have no support for

these figures, but at present there is at least no satisfactory evidence in support of any others.

Our own observations on stillbirths bring us no nearer to an acceptable conception ratio; indeed the difficulty of collecting the requisite evidence seems if anything more formidable than ever. Certainly we can have no confidence about sex ratios of abortions in the early months; nor can we argue from values in the late months of pregnancy, since it is quite certain that different causes of abortion have different sex ratios, and more than likely that their relative proportions are changing from month to month. Moreover, as the sex ratios of certain stillborn congenital malformations are different at the seventh, eighth and ninth months, we can no longer assume that the sex ratio of any single cause of abortion remains constant throughout pregnancy. We must therefore rely upon examination of representative specimens, and here we are confronted not only with the difficulties to which we have referred, but also by the possibility that fertilization may be followed by abortion with little interruption of the cycle (Hertig and Rock, 1949). In these circumstances no estimate of the sex ratio at conception seems warranted until we have much fuller information for different stages of gestation on:

(1) the pathology and incidence of abortions;

(2) the sex ratio of different causes of abortion; and

(3) the reason for differences in the sex ratios of different causes of abortion.

SUMMARY

1. Records for Birmingham, England, for the years 1936-1949 have been used in an investigation of the sex ratio of stillbirths (foetuses born dead after the 28th week) by cause and duration of gestation. In this period there were 7,066 stillbirths from single pregnancies, and they were grouped by cause of death in four classes which correspond roughly with those used in the Annual Reports of the Registrar-General for Scotland.

2. The sex ratio of stillbirths increases as the duration of gestation increases (from 49.8 at the seventh month to 57.1 at nine months and over). This increase is explained by:

(a) different sex ratios of the four classes of stillbirth;

(b) changes in the proportion of stillbirths attributed to each class;

(c) an increase with duration of gestation in the sex ratio of the class "foetal malformation."

3. The increase in the sex ratio of the class "foetal malformation" is explained by:

(a) an increase with duration of gestation in the sex ratios of anencephalus and spina bifida (mainly); and

(b) changes in the proportion of stillbirths attributed to each malformation.

4. For every cause of death, the sex ratio is higher for macerated than for not macerated stillbirths. The sex ratio of macerated foetuses increases with duration of gestation. This presumably means that after death males are retained in the uterus longer than females.

These observations are discussed in relation to the sex of abortions earlier than the 28th week, and to previous estimates of the sex ratio at conception.

REFERENCES

AUERBACH, E. (1912). Das wahre Geschlechtsverhältnis des Menschen. *Arch. Rass.-u. Gesbiol.*, 9: 10.

BAYER, R. (1939). Zur Frage der Übergeburtlichkeit männlicher Früchte in verschiedenen Fetalmonaten. *Arch. Gynäk.*, 169: 619.

BERTILLON, J. (1893). De la mortalité par âge avant la naissance. *J. Soc. Statist. Paris*, 34: 167.

BOLDRINI, M. (1930). Sulla proporzione dei sessi nei concepimenti e nelle nascite. Univ. Catholica del Sacro Cuore., *Contrib. del Lab. di Statist.*, 1: 213.

BRUNNER, E. K. (1941). The outcome of 1556 conceptions. *Human Biol.*, 13: 159.

BRUNNER, E. K., and L. NEWTON. (1939). Abortions in relation to viable births in 10,609 pregnancies. *Amer. J. Obstet. Gynec.*, 38: 82.

CARVALLO, M. E. (1912). La masculinité dans les naissances humaines. *C. R. Ass. franç. Av. Sci.*, 41: 145.

CIOCCO, A. (1938). The masculinity of stillbirths and abortions in relation to duration of uterogestation and to the stated causes of foetal mortality. *Human Biol.*, 10: 235.

HERTIG, A. T. (1943). Minimal criteria required to prove prima facie case of traumatic abortion or miscarriage. *Ann. Surg.*, 117: 596.

HERTIG, A. T. and J. ROCK. (1949). A series of potentially abortive ova recovered from fertile women prior to the first missed period. *Amer. J. Obstet. Gynec.*, 58: 968.

LOWE, C. R., and T. MCKEOWN. (1950). The sex ratio of human births related to maternal age. *Brit. J. Soc. Med.*, 4: 75.

NICHOLS, J. B. (1907). The numerical proportion of the sexes at birth. *Mem. Amer. Anthrop. Ass.*, 1: 247.

PARKES, A. S. (1926). The mammalian sex ratio. *Biol. Rev.*, 1: 1.

RECORD, R. G., and T. MCKEOWN. (1949). Congenital malformations of the central nervous system. *Brit. J. Soc. Med.*, 3: 183.

SCHULTZ, A. H. (1921). Sex incidence in abortions. *Contr. Embryol. Carneg. Instn.*, 12: 177.

SPAULDING, M. H. (1921). The development of the external genitalia of the human embryo. *Contr. Embryol. Carneg. Instn.*, 13: 69.

STREETER, G. L. (1920). Weight, sitting height, head size, foot length and menstrual age of the human embryo. *Contr. Embryol. Carneg. Instn.*, 11: 143.

TIETZE, C. (1948). A note on the sex ratio of abortions. *Human Biol.*, 20: 156.

TSCHUPROW, AL. A. (1915). Zur Frage des sinkenden Knaubenüberschusses unter den ebelich Gebornen. *Bull. Inst. int. Statist.*, 20: 378.

WIEHL, D. G. (1938). A summary of data on the reported incidence of abortion. *Milbank Mem. Fd. Quart. Bull.*, 16: 1.

WILSON, K. M. (1926). Correlation of external genitalia and sex glands in the human embryo. *Contrib. Embryol. Carneg. Instn.*, 18: 23.

APPENDIX

Sex ratio of stillbirths related to cause and duration of gestation, Birmingham, 1936-49

STILLBIRTH ATTRIBUTED TO:	DURATION OF GESTATION (months)				
	7	8	9 and over	Unspecified	Total
1. Disease in or accident to the mother					
(a) Not specially associated with pregnancy	61.3 (75)	57.0 (79)	60.6 (109)	41.7 (12)	58.9±3.0 (275)
(b) Toxaemia	50.9 (216)	50.7 (292)	53.6 (278)	48.9 (45)	51.6±1.7 (831)
(c) Toxaemia with ante-partum haemorrhage	55.9 (93)	57.0 (107)	60.7 (107)	75.0 (20)	59.0±2.7 (327)
(d) Placenta praevia	54.7 (53)	53.3 (75)	49.2 (63)	50.0 (8)	52.3±3.5 (199)
(e) Unexplained or unspecified ante-partum haemorrhage	65.4 (78)	60.3 (78)	58.3 (103)	50.0 (12)	60.5±3.0 (271)

APPENDIX (Continued)

Sex ratio of stillbirths related to cause and duration of gestation, Birmingham, 1936-49

STILLBIRTH ATTRIBUTED TO:	DURATION OF GESTATION (months)				
	7	8	9 and over	Unspecified	Total
2. Foetal malformation					
(a) Anencephalus alone	23.4 (137)	31.4 (204)	38.7 (137)	21.4 (14)	30.9±2.1 (492)
(b) Anencephalus with spina bifida	17.1 (35)	30.8 (26)	37.5 (8)	...	24.6±5.2 (69)
(c) Hydrocephalus alone	61.1 (18)	46.5 (43)	53.5 (127)	71.4 (7)	53.3±3.6 (195)
(d) Hydrocephalus with spina bifida	40.0 (5)	26.3 (19)	32.1 (81)	66.7 (3)	32.4±4.5 (108)
(e) Spina bifida alone	16.7 (18)	30.4 (23)	44.7 (38)	40.0 (5)	34.5±5.1 (84)
(f) Other malformations	51.3 (39)	61.0 (41)	43.6 (78)	44.4 (9)	49.7±3.9 (167)
3. Difficult labour					
(a) Breech delivery	57.1 (56)	62.5 (88)	58.6 (553)	64.9 (37)	59.3±1.8 (734)
(b) Prolapsed cord (other than breech)	83.3 (6)	55.0 (20)	65.2 (184)	81.3 (16)	65.9±3.2 (226)
(c) Pressure on cord	0 (1)	66.7 (15)	56.6 (205)	60.0 (10)	57.5±3.3 (231)
(d) Other difficult labour	50.0 (14)	64.0 (50)	65.8 (819)	53.2 (47)	64.8±1.6 (930)
4. Ill-defined or unknown cause	53.4 (395)	55.6 (399)	55.0 (1003)	56.2 (130)	54.9±1.1 (1927)
Totals	49.8±1.4 (1239)	51.3±1.3 (1559)	57.1±0.8 (3893)	55.7±2.6 (375)	54.5±0.6 (7066)

ACCUMULATION OF NITROGEN AND SIX MINERALS IN THE HUMAN FETUS DURING GESTATION

BY HARRIET J. KELLY, RALPH E. SLOAN, WALTER HOFFMAN, AND CLARIBEL SAUNDERS

Research Laboratory, Children's Fund of Michigan, Detroit

THE composition of the human fetus at various stages in gestation is extremely significant in any study of metabolism and growth during the reproductive cycle. In searching for the best available values representing the amounts of chemical constituents in fetuses, relatively few analytical results were found and compilation of the data was hampered by discrepancies in methods of calculation and presentation. For these reasons it seemed that collection of data from the greatest possible number of reports giving comparable results from analyses of human fetuses, followed by calculation of curves for the contents of the different chemical components in relation to fetal age, would constitute a valuable contribution to the literature.

The most recent report available of chemical analyses of individual human fetuses was made by Iob and Swanson (1) in 1934 and presented results of analyses of 16 fetuses. A year earlier Givens and Macy (2) had published analyses of 25 fetuses and assembled values from the literature for 71 others. A search of the literature did not reveal other data on the composition of individual human fetuses in American journals, although Widdowson (3) recently published average data from analyses of six "full term" human fetuses.

In making a routine check of the references and data assembled in the latest papers (1, 2), inspection of the early papers was not sufficient to clearly distinguish between data derived from analyses of fetal tissue and those from determinations of the composition of human milk. In some of the earlier papers, also, it was difficult to be sure whether fetuses analyzed were human or of another species. In addition, data were presented in a variety of degrees of completeness and in several languages.

These facts, plus recognition of possible errors arising from repeated quotation of values over a period of many years, indicated the need for complete and thorough translation of the early papers.

Derivation of curves indicative of the accumulation of chemical components in the human fetus during gestation is contingent upon comparable estimates of fetal age. In some of the early papers the fetuses analyzed were "judged" or "seemed" to be "about" a certain age. In other papers various criteria were used to obtain a designation of fetal age. Evidence indicates that the most reliable estimates are based upon the crown-heel length of the body and Scammon and Calkins (4) published data from which they derived a formula which is applicable to fetuses of all sizes. Unfortunately, of the papers available presenting analyses of human fetuses, only eleven contained records of the lengths of the fetuses analyzed. Using the lengths given by the authors the age of each fetus described in the eleven papers was calculated (4) in lunar months from date of last menstruation and the ages so calculated were used in deriving curves for the accumulation of nitrogen, calcium, magnesium, potassium, sodium, phosphorus, and chlorine in the fetus during gestation.

RESULTS

In Table 1 all fetuses are listed and numbered in order of their calculated ages with the authors' data for body length, wet weight, dry weight, fat and ash contents, sex, and reference number referring to the list appended to this paper. Table 2 presents the author's data for the nitrogen, calcium, magnesium, sodium, potassium, phosphorus, and chlorine contents of each fetus and values per kilogram of body weight and per 100 gm. of ash. When these values were not given by the authors they were calculated with the data provided and checked thoroughly. Data in Table 2 are given with the assigned subject numbers and the calculated ages of the fetuses, eliminating any confusion in comparing values for the same fetus.

The use of data such as are given in Table 2 presents numerous difficulties. The values represent unknown variations in methods of collecting, preparation and analysis, superimposed upon possible racial, nutritional, and medical differences in subjects. Inspection of the data shows that for any fetal age more than one mean could be derived by altering the basis upon which the values were separated into the groups

TABLE 1

Length, age and weights of fetuses for which comparable analytical values are available*

SUBJECT NUMBER	AGE (lunar months)	BODY LENGTH (cm.)	BODY WEIGHT (gm.)				SEX	REFERENCE NUMBER
			Wet	Dry	Fat	Ash		
1	2.5	2.5	0.975	0.24		0.001	F	5
2	2.6	3.0	1.7	0.1				1
Mean	2.5	2.8	1.3	0.17		0.001		
3	2.8	5.0	3.91	0.252		0.0243		6
4	2.9	5.9	6.8	0.487		0.053		6
5	3.0	7.4	8.6	0.524		0.0414		6
6	3.1	8.0	12.9	1.01		0.104		6
7	3.1	8.0	18.2	1.1				1
8	3.2	8.5	12.6	1.045		0.101		6
9	3.2	8.7	13.7	1.146		0.115		6
10†	3.2	9.0	6.4	0.49		0.02	M	2
Mean	3.0	7.4	11.0	0.79		0.073		
11	3.3	9.5	23.9	1.4				1
12	3.3	9.5	14.0	1.47		0.15	M	2
13	3.3	9.6	11.0	1.26		0.11	M	2
14	3.3	9.8	18.9	1.671		0.184		6
15	3.4	10.6	19.0	2.27		0.20	M	2
16	3.4	10.6	26.75	2.966		0.342		6
17	3.4	11.0	33.0	3.04		0.32	M	2
18	3.5	11.5	32.0	4.40		0.60	M	2
19	3.5	11.5	42.4	3.658		0.392		6
20	3.5	11.6	34.48	2.713		0.248		6
21	3.6	12.0	36.5	3.0	0.21	0.358	M	5
22	3.6	12.5	33.0	3.31		0.48	M	2
23	3.7	13.0	41.6	3.687		0.479		6
24	3.7	13.1	131.0	15.28		1.60	M	2
25	3.7	13.5	56.5	5.1	0.25	0.571	M	5
26	3.7	13.5	59.4	4.5	0.30	0.81		1
Mean	3.5	11.4	38.3	3.73	0.25	0.456		
27	3.8	13.6	72.5	5.89		0.696		6
28	3.9	15.0	66.0	6.081				6
29	3.9	15.0	72.6	6.89		0.839		6
30	3.9	15.0	71.0	7.16		1.02	M	2
31	4.0	16.0	62.35	6.578		0.799		6
32	4.0	16.0	86.0	8.90		1.02	M	2
33	4.2	17.5	114.5	12.9	0.80	1.32		1
Mean	4.0	15.4	77.8	7.77	0.80	0.949		

TABLE 1 (Continued)

Length, age and weights of fetuses for which comparable analytical values are available*

SUBJECT NUMBER	AGE (lunar months)	BODY LENGTH (cm.)	BODY WEIGHT (gm.)				SEX	REFERENCE NUMBER
			Wet	Dry	Fat	Ash		
34	4.3	18.0	111					7
35	4.4	18.3	128.0	14.28		1.43	M	2
36	4.4	18.5	95.5	8.9	0.46	1.337	M	5
37	4.4	18.5	104.7	8.2	0.53	1.089	M	5
38	4.4	19.0	156.8	14.6	0.85	2.242	F	5
39	4.4	19.0	122.2	13.3		1.779		6
40	4.6	20.0	201.0	16.70		1.87	M	2
41	4.7	21.0	113.0	18.12		2.83	F	2
42	4.7	21.0	161.0	19.52		2.66	M	2
43	4.7	21.0	210.0	25.50		2.93	M	2
Mean	4.5	19.4	140.3	15.46	0.61	2.018		
44	4.8	21.5	244.0	22.0	0.68	2.830	M	5
45	4.9	22.0	172.0	26.74		3.54	F	2
46	5.0	22.5	235.5	24.0	1.34	3.862	M	5
47	5.0	23.0	249.1	31.38		4.447		6
48	5.0	23.0	264.0	29.5	1.37	4.990	F	5
49	5.1	23.5	259.0	32.8	2.07	3.99		1
50	5.1	23.5	360.0	41.13		4.84	M	2
51	5.1	23.5	248					8
52	5.2	24.0	244.5	28.334		3.995		6
53	5.2	24.0	299.0	32.8	1.79	5.711	F	5
54	5.2	24.0	312.0	37.73		4.92	F	2
55	5.2	24.1	296.0	36.38		5.15	M	2
56	5.2	24.5	373.0	38.169		5.176		6
Mean	5.0	23.3	273.5	31.75	1.45	4.454		
57	5.4	25.4	335.0	41.1	2.34	6.97		1
58	5.5	26.0	361.8	39.1	2.60	7.019	M	5
59	5.7	27.3	490.0	60.5	3.43	12.94		1
Mean	5.5	26.2	395.6	46.90	2.79	8.976		
60	5.8	28.0	433.0	54.38		8.57	M	2
61	6.0	29.5	314.0	65.89		10.27	F	2
62	6.1	30.0	575.0	79.5	6.10	13.398	F	5
63	6.2	30.5	590.0	85.7	7.08	15.81		1
Mean	6.0	29.5	478.0	71.37	6.59	12.012		
64	6.4	31.5	570.0	82.7	7.41	13.51		1
65	6.5	32.0	584.0	80.0		11.702		6
66	6.7	33.5	771.0	125.2	15.26	21.896	M	5
Mean	6.5	32.3	641.7	95.97	11.34	15.703		

TABLE 1 (Continued)

Length, age and weights of fetuses for which comparable analytical values are available*

SUBJECT NUMBER	AGE (lunar months)	BODY LENGTH (cm.)	BODY WEIGHT (gm.)				SEX	REFERENCE NUMBER
			Wet	Dry	Fat	Ash		
67	6.8	34.0	832.9	138.2	22.49	18.990	M	5
68	6.9	34.5	910.0	159.0	31.58	26.754	F	5
69	7.0	35.0	1117.0	170.71	26.36	28.372	M	5
70	7.0	35.0	841.0					8
71	7.2	36.0	836.0	136.5	18.48	23.826	F	5
72	7.2	36.0	956.3	196.41	48.02	26.1		9
73	7.2	36.1	1010.0	146.3	22.22	21.92		1
74	7.2	36.5	1071.0	161.37		18.69	M	2
75	7.2	36.5	1107.0	177.62		35.85	F	2
Mean	7.0	35.5	964.6	160.76	28.19	25.063		
76	7.5	38.0	928.0	159.5	22.64	26.170	M	5
77	7.6	38.5	960.0	162.1	21.12	26.02		1
78	7.6	38.6	1205.0	211.4	42.18	28.92		1
Mean	7.5	38.4	1031.0	177.67	28.65	27.037		
79†	7.8	39.5	1060.0	199.0		17.59	F	2
80†	7.9	40.0	1170.0	266.5		42.64	M	2
81	7.9	40.0	1339					8
82	8.2	41.3	1555.0	316.9	74.64	37.79		1
83	8.2	41.7	1545.0	275.4	52.53	39.71		1
Mean	8.0	41.0	1479.7	296.15	63.58	38.750		
84	8.4	42.5	1615.0	306.6	62.98	39.89		1
85	8.7	44.0	1760.6	456.1	153.17	58.100	M	5
Mean	8.5	43.2	1687.8	381.35	108.08	48.995		
86	8.9	45.0	1495.7	391.2	76.43	31.559	M	5
Mean	9.0	45.0	1495.7	391.20	76.43	31.559		
87	9.3	47.0	1875	463.5	157.9	58.5	M	10
88	9.4	47.5	2476	669	270	76	F	11
89	9.7	49.0	2616	742	358	54	F	12
Mean	9.5	47.8	2322.3	624.83	261.97	62.833		
90	9.8	49.5	2683	725	273	72	M	11
91	9.8	49.5	2755	850	443	74	M	12
92	10.0	50.0	3048	854	366	85	F	13
93	10.3	51.8	2915.0	714.0	195.3	91.82		1
94	10.5	52.5	3348	930	378	87	M	13
95	10.7	53.5	3294.8	855.52	299.83	84.017	M	5
Mean	10.0	51.1	3007.3	821.42	325.86	82.306		

* Ages were calculated from body lengths with the formula (4):

$$\text{age} = 2.3 + \frac{2.5}{28}(\text{length}) + \frac{(\text{length})^2}{784}.$$

† Data for fetus number 10, which were at variance with other values, and for numbers 79 and 80, twins, were not included with the data compiled in Table 2.

TABLE 2

Values reported for constituents of 92 human fetuses

Values in grams.

SUBJECT NUMBER	AGE (lunar months)	CALCIUM		
		Per fetus	Per kilogram body weight	Per 100 gm. ash
3	2.8	0.0034	0.86	14.0
4	2.9	0.0080	1.14	15.1
5	3.0	0.0079	0.91	19.2
6	3.1	0.020	1.55	19.2
7	3.1	0.016	0.97	
8	3.2	0.019	1.50	18.7
11	3.3	0.040	1.74	
12	3.3	0.041	2.93	27.1
13	3.3	0.034	3.09	31.2
14	3.3	0.041	2.19	22.5
15	3.4	0.047	2.47	23.5
16	3.4	0.072	2.69	21.1
17	3.4	0.071	2.15	22.3
18	3.5	0.162	5.06	27.0
19	3.5	0.089	2.07	22.6
20	3.5	0.054	1.59	21.6
22	3.6	0.083	2.52	17.3
23	3.7	0.12	2.79	24.5
24	3.7	0.47	3.56	29.3
26	3.7	0.11	1.91	13.8
27	3.8	0.14	1.93	20.5
28	3.9	0.18	2.72	
29	3.9	0.17	2.29	20.1
30	3.9	0.26	3.62	25.3
31	4.0	0.15	2.36	18.9
32	4.0	0.24	2.83	23.9
33	4.2	0.36	3.18	27.6
35	4.4	0.37	2.92	26.2
40	4.6	0.58	2.88	30.9
41	4.7	0.84	7.45	29.7
42	4.7	0.79	4.89	29.6
43	4.7	0.84	4.02	28.8
45	4.9	1.04	6.04	29.3
47	5.0	1.23	4.93	27.9
49	5.1	1.26	4.88	31.6
50	5.1	1.42	3.93	29.2
52	5.2	1.05	4.29	26.2
54	5.2	1.48	4.76	30.2
55	5.2	1.58	5.33	30.7

TABLE 2 (Continued)

Values reported for constituents of 92 human fetuses

Values in grams.

SUBJECT NUMBER	AGE (lunar months)	CALCIUM		
		Per fetus	Per kilogram body weight	Per 100 gm. ash
56	5.2	1.43	3.79	28.0
57	5.4	1.58	4.71	22.6
59	5.7	3.27	6.68	25.3
60	5.8	2.97	6.87	34.7
61	6.0	3.67	11.69	35.7
63	6.2	3.88	6.58	24.6
64	6.4	3.21	5.64	23.8
65	6.5	3.33	5.70	28.5
72	7.2	7.42	7.76	28.4
73	7.2	5.01	4.95	22.9
74	7.2	5.98	5.58	32.0
75	7.2	11.51	10.40	32.1
77	7.6	6.78	7.06	26.0
78	7.6	6.81	5.66	23.5
82	8.2	9.36	6.01	24.8
83	8.2	9.04	5.86	22.8
84	8.4	9.62	5.96	24.1
87	9.3	15.14	8.08	26.0
88	9.4	22.94	9.26	30.2
89	9.7	13.08	5.00	24.2
90	9.8	17.94	6.69	24.9
91	9.8	20.37	7.43	27.5
92	10.0	23.23	7.62	27.3
93	10.3	24.38	8.36	26.5
94	10.5	25.01	7.47	28.8
		MAGNESIUM		
12	3.3	0.004	0.29	2.7
13	3.3	0.003	0.27	2.8
15	3.4	0.003	0.16	1.6
17	3.4	0.007	0.21	2.1
18	3.5	0.033	1.03	5.6
22	3.6	0.045	1.36	9.3
24	3.7	0.036	0.28	2.2
26	3.7	0.007	0.11	0.9
30	3.9	0.082	1.16	8.0
32	4.0	0.058	0.67	5.7
33	4.2	0.017	0.15	1.3

TABLE 2 (Continued)

Values reported for constituents of 92 human fetuses

Values in grams.

SUBJECT NUMBER	AGE (lunar months)	MAGNESIUM		
		Per fetus	Per kilogram body weight	Per 100 gm. ash
35	4.4	0.030	0.23	2.1
40	4.6	0.066	0.33	3.5
41	4.7	0.087	0.77	3.1
42	4.7	0.057	0.35	2.4
43	4.7	0.023	0.11	0.8
45	4.9	0.031	0.18	0.9
49	5.1	0.049	0.18	1.2
50	5.1	0.113	0.31	2.3
54	5.2	0.077	0.25	1.6
55	5.2	0.170	0.57	3.3
57	5.4	0.063	0.19	0.9
59	5.7	0.102	0.21	0.8
60	5.8	0.282	0.6[illegible]	3.3
61	6.0	0.303	0.96	2.9
63	6.2	0.104	0.18	0.6
64	6.4	0.112	0.20	0.8
72	7.2	0.189	0.20	0.7
73	7.2	0.139	0.14	0.6
74	7.2	0.324	0.30	1.7
75	7.2	0.706	0.64	2.0
77	7.6	0.199	0.21	0.8
78	7.6	0.243	0.20	0.8
82	8.2	0.350	0.22	0.9
83	8.2	0.365	0.24	0.9
84	8.4	0.314	0.19	0.8
87	9.3	0.452	0.24	0.9
88	9.4	0.513	0.21	0.7
89	9.7	0.434	0.18	0.8
90	9.8	0.603	0.22	0.8
91	9.8	0.277	0.12	0.4
92	10.0	0.501	0.16	0.6
93	10.3	0.783	0.27	0.8
94	10.5	0.482	0.14	0.6
		SODIUM		
7	3.1	0.021	1.150	
11	3.3	0.018	0.793	
26	3.7	0.103	1.757	12.72

TABLE 2 (Continued)

Values reported for constituents of 92 human fetuses

Values in grams.

SUBJECT NUMBER	AGE (lunar months)	SODIUM		
		Per fetus	Per kilogram body weight	Per 100 gm. ash
33	4.2	0.267	2.325	20.23
49	5.1	0.619	2.385	15.51
57	5.4	0.720	2.171	10.33
59	5.7	1.161	2.376	8.97
63	6.2	1.286	2.192	8.13
64	6.4	1.230	2.171	9.10
72	7.2	2.389	2.498	9.16
73	7.2	1.925	1.909	8.78
77	7.6	2.263	2.357	8.70
78	7.6	2.410	2.014	8.33
82	8.2	2.925	1.886	7.74
83	8.2	3.128	2.035	7.88
84	8.4	3.295	2.051	8.26
88	9.4	4.600	1.858	6.05
89	9.7	4.051	1.558	7.50
90	9.8	4.125	1.537	5.73
91	9.8	4.518	1.632	6.10
92	10.0	5.535	1.816	6.51
93	10.3	5.528	1.752	6.02
94	10.5	5.943	1.775	6.83
		POTASSIUM		
7	3.1	0.008	0.340	
11	3.3	0.008	0.320	
26	3.7	0.066	1.087	8.15
33	4.2	0.192	1.669	14.54
49	5.1	0.438	1.685	10.98
57	5.4	0.512	1.388	7.34
59	5.7	0.704	1.439	5.44
63	6.2	0.915	1.548	5.79
64	6.4	0.876	1.188	6.48
72	7.2	1.644	1.719	6.31
73	7.2	1.525	1.505	6.96
77	7.6	1.400	1.454	5.38
78	7.6	2.201	1.830	7.61
82	8.2	2.608	1.681	6.90
83	8.2	2.619	1.697	6.60
84	8.4	2.803	1.740	7.03

TABLE 2 (Continued)

Values reported for constituents of 92 human fetuses

Values in grams.

SUBJECT NUMBER	AGE (lunar months)	POTASSIUM		
		Per fetus	Per kilogram body weight	Per 100 gm. ash
88	9.4	3.528	1.425	4.64
89	9.7	4.001	1.494	7.41
90	9.8	4.516	1.683	6.27
91	9.8	4.151	1.494	5.61
92	10.0	5.105	1.675	6.00
93	10.3	4.903	1.689	5.34
94	10.5	5.562	1.661	6.39
		PHOSPHORUS		
7	3.1	0.015	0.94	
11	3.3	0.028	1.23	
26	3.7	0.102	1.68	12.59
33	4.2	0.245	2.15	18.56
49	5.1	0.889	3.44	22.28
57	5.4	1.097	3.26	15.74
59	5.7	2.045	4.15	15.80
63	6.2	2.429	4.10	15.36
64	6.4	2.184	3.81	16.16
72	7.2	3.557	3.72	13.63
73	7.2	3.442	3.39	15.70
77	7.6	4.043	4.21	15.54
78	7.6	4.579	3.78	15.83
82	8.2	6.301	4.03	16.67
83	8.2	6.218	4.01	15.66
84	8.4	6.555	4.04	16.43
87	9.3	9.411	5.02	16.05
88	9.4	13.094	5.29	17.23
89	9.7	8.947	3.40	16.57
90	9.8	12.177	4.54	16.91
91	9.8	12.919	4.67	17.46
92	10.0	13.705	4.50	16.12
93	10.3	14.957	5.11	16.29
94	10.5	14.578	4.35	16.76
		CHLORINE		
11	3.3	0.018	0.75	
26	3.7	0.12	2.00	14.81

TABLE 2 (Continued)

Values reported for constituents of 92 human fetuses

Values in grams.

SUBJECT NUMBER	AGE (lunar months)	CHLORINE		
		Per fetus	Per kilogram body weight	Per 100 gm. ash
33	4.2	0.31	2.73	23.33
34	4.3	0.28	2.52	
49	5.1	0.70	2.70	17.49
51	5.1	0.68	2.73	
57	5.4	0.88	2.62	12.61
59	5.7	1.20	2.45	9.29
63	6.2	1.46	2.48	9.24
64	6.4	1.45	2.55	10.73
70	7.0	1.87	2.23	
72	7.2	2.71	2.83	10.38
73	7.2	2.23	2.20	10.17
77	7.6	2.27	2.18	8.72
78	7.6	2.88	2.39	9.97
81	7.9	2.96	2.21	
82	8.2	3.54	2.27	9.35
83	8.2	3.52	2.28	8.87
84	8.4	3.62	2.24	9.07
88	9.4	4.47	1.8	5.88
89	9.7	4.80	1.8	8.89
90	9.8	4.07	1.5	5.65
91	9.8	4.88	1.8	6.59
92	10.0	5.68	1.9	6.68
93	10.3	5.29	1.8	5.76
94	10.5	5.97	1.8	6.86

		NITROGEN	
		Per fetus	Per kg.
2	2.6	0.01	5.7
11	3.3	0.16	8.6
26	3.7	0.46	7.8
33	4.2	1.35	11.8
49	5.1	3.46	13.4
57	5.4	4.29	12.7
59	5.7	6.12	12.5
63	6.2	8.77	14.9
64	6.4	8.49	14.9
72	7.2	16.77	17.4

TABLE 2 (Continued)

Values reported for constituents of 92 human fetuses

Values in grams.

SUBJECT NUMBER	AGE (lunar months)	NITROGEN	
		Per fetus	Per kg.
73	7.2	14.60	14.4
77	7.6	16.38	17.1
78	7.6	18.50	15.4
82	8.2	26.59	17.1
83	8.2	25.37	16.4
84	8.4	28.06	17.4
88	9.4	49.0	19.8
89	9.7	46.8	17.9
90	9.8	58.5	21.8
91	9.8	50.5	18.3
92	10.0	59.9	19.6
93	10.3	57.5	19.7
94	10.5	70.2	21.0

to be averaged. Also, since the numbers of determinations vary for the different chemical elements, the best division into age groups for one element might not be the best grouping for all other elements. For this reason expenditure of the time required to obtain curves fitted to the data seemed justified.

By the method of least squares a satisfactory "fit" was obtained with second-degree curves for magnesium, sodium, potassium, phosphorus, and for chlorine. A satisfactory calcium curve was obtained by using orthogonal polynomials (14). For nitrogen the best fit was provided by a straight line fitted by the method of least squares to the data per kilogram of body weight. The equations for the curves derived are:

$$\text{Mg (gm.)} = .00436\ (\text{age})^2 + .020012\ (\text{age}) - .1111,$$
$$\text{Cl (gm.)} = .05984\ (\text{age})^2 - .009490\ (\text{age}) - .7285,$$
$$\text{K (gm.)} = .10558\ (\text{age})^2 - .718291\ (\text{age}) + 1.278,$$
$$\text{P (gm.)} = .35014\ (\text{age})^2 - 2.764922\ (\text{age}) + 5.567,$$
$$\text{Na (gm.)} = .08176\ (\text{age})^2 - .3431\ (\text{age}) + .2732,$$

$$N \text{ (gm./kg. body weight)} = 1.6861 \text{ (age)} + 3.2576,$$

$$\begin{aligned} Ca \text{ (gm.)} = {} & .0044 \text{ (age} - 6.4)^5 + .0274 \text{ (age} - 6.4)^4 \\ & + .00936 \text{ (age} - 6.4)^3 + .2173 \text{(age} - 6.4)^2 \\ & + 2.284 \text{ (age} - 6.4) + 3.9963. \end{aligned}$$

Using the equations, the amounts of nitrogen and the six minerals in the human fetus were calculated for bi-weekly intervals for the fourth through the tenth lunar month and are given in Table 3. With

TABLE 3

*Constituents of the human fetus at age intervals in gestation as indicated by analyses of 92 fetuses**

Values in grams per fetus.

AGE (lunar months)	CALCIUM	MAGNESIUM	CHLORINE	SODIUM	POTASSIUM	PHOSPHORUS	NITROGEN
4.0	0.195	0.039	0.191	0.209	0.094	0.109	0.855
4.5	0.624	0.067	0.440	0.385	0.184	0.215	1.777
5.0	1.281	0.098	0.720	0.602	0.326	0.496	3.153
5.5	2.124	0.131	1.029	0.860	0.521	0.952	4.792
6.0	3.118	0.166	1.369	1.158	0.769	1.582	6.755
6.5	4.226	0.203	1.738	1.498	1.070	2.388	9.878
7.0	5.451	0.243	2.137	1.878	1.424	3.369	13.239
7.5	6.830	0.284	2.566	2.299	1.830	4.525	18.422
8.0	8.471	0.328	3.025	2.761	2.289	5.856	23.436
8.5	10.551	0.374	3.514	3.264	2.801	7.362	27.340
9.0	13.343	0.422	4.033	3.808	3.365	9.044	33.828
9.5	17.234	0.473	4.582	4.393	3.983	10.900	43.855
10.0	22.734	0.525	5.160	5.019	4.653	12.931	55.908

* Data derived from curves fitted to values assembled (Table 2) for each element.

† Based upon application of the equation and conversion from per kilogram of body weight to per fetus with mean values (Table 1) smoothed by three-point averaging.

the few samples available the method could not be extended to obtain values in the first three lunar months of gestation, the period in which the body is small, making the variations in the measurements and chemical determinations relatively larger.

SUMMARY

Eleven papers presenting analyses of 95 human fetuses provided data permitting calculation of fetal age from the crown-heel lengths of the bodies.

Using the calculated fetal ages, curves were derived for the fetal contents of nitrogen, calcium, magnesium, sodium, potassium, phosphorus and chlorine during the period from four through ten months from date of last menstruation.

Mean values are given which facilitate conversion of the data per fetus to the bases of body length and weight and content of solids, ash, and fat.

REFERENCES

1. Iob, V., and W. W. Swanson. Mineral growth of the human fetus. *Am. J. Diseases Children*, 47: 302, 1934.
2. Givens, M. H., and I. G. Macy. The chemical composition of the human fetus. *J. Biol. Chem.*, 102: 7, 1933.
3. Widdowson, E. M. Chemical composition of newly born mammals. *Nature*, 166: 626, 1950.
4. Scammon, R. E., and L. A. Calkins. The Development and Growth of the External Dimensions of the Human Body in the Fetal Period. *Minneapolis*: Univ. Minnesota Press, 1929.
5. Fehling, H. Beiträge zur Physiologie des placentaren Stoffverkehrs. *Arch. Gynäkol.*, 11: 523, 1877.
6. Schmitz, E. Untersuchungen über den Kalkgehalt der wachsenden Frucht. *Arch. Gynäkol.*, 121: 1, 1923.
7. Rosemann, R. Über den Gesamtchlorgehalt des menschlichen Fötus. *Pflüger's Arch. Physiol.*, 142: 459, 1911.
8. Rosemann, R. Über den Gesamtchlorgehalt des tierischen Körpers. *Pflüger's Arch. Physiol.*, 135: 177, 1910.
9. Langstein, L., and F. Edelstein. Die chemische Zusammensetzung frühgeborener Säuglinge und ihr Wachstumsansatz. *Z. Kinderheik.*, 15: 49, 1917.
10. Brubacher, H. Ueber den Gehalt an anorganischen Stoffen, besonders an Kalk, in den Knochen und Organen normaler und rhachitischer Kinder. *Z. Biol.*, Neue Folge 9: 517, 1891.
11. Camerer, W., Jr. Die chemische Zusammensetzung des Neugeborenen. *Z. Biol.*, 40: 529, 1900.
12. Camerer, W., Jr. Die chemische Zusammensetzung des Neugeborenen. *Z. Biol.*, 39: 173, 1899-1900.
13. Camerer, W., Jr. Die chemische Zusammensetzung des neugeborenen Menschen. *Z. Biol.*, 43: 1, 1902.
14. Anderson, R. L., and E. E. Houseman. Tables of Orthogonal Polynomial Values Extended to N = 104. Iowa State Coll. Agr. Mech. Arts, *Agr. Exp. Sta. Research Bull. No. 297*, 1942.

CHARLES PAINE WINSOR

JUNE 19, 1895 — APRIL 4, 1951

CHARLES PAINE WINSOR
EDITOR
of
HUMAN BIOLOGY

IT is with deep regret that we announce the sudden death on April 4 of Dr. Charles Paine Winsor, Editor of Human Biology. He is survived by his wife, Agnes Winsor.

Dr. Winsor was born in Boston in 1895. He took his A.B. and S.B. degrees in Engineering at Harvard and worked from 1921 to 1927 as an engineer with the New England Telephone and Telegraph Company. An interest in biology brought him to Baltimore to work as assistant to Raymond Pearl. After some years with Dr. Pearl he returned to Harvard to complete his Ph.D. degree in general physiology in 1935 under W. J. Crozier. From 1938 to 1941 he worked in the Statistical Laboratory at Iowa State College as Assistant Professor of Mathematics, and during the war he was engaged in research at Princeton under a contract with the Office of Scientific Research and Development. In 1946 he returned to Baltimore as Assistant Professor of Biostatistics in the School of Hygiene and Public Health, Johns Hopkins University, at which time he assumed the editorship of Human Biology.

Dr. Winsor was endowed with a vast fund of knowledge of biology, statistics and human affairs, and with exceptional powers of reasoning. His research writings were highly respected for their characteristic clarity in attacking a complex problem, and for their keen sense of what was important and what was unimportant. His counsel and criticism were valued by the leading scientists in his field. He greatly relished his work as editor and would spare no effort in helping authors towards a sound and lucid exposition of their findings.

One could not wish for a more delightful or stimulating companion. He excelled in the art of conversation; on almost any subject he had a striking comment, usually unorthodox, or an illuminating anecdote. His love of heated debate, which can create enemies for a man, brought him instead a widening range of friendships, for his opponents were always disarmed by his modesty and good humor, and his willingness to turn around and argue against himself. His death is a great loss to biometry and a keen personal loss to his many friends.

WILLIAM G. COCHRAN